Applications of Optimal Transport to Spatial Signal Processing Problems

In Defense of a Dissertation

David Gentile, 16 April 2026

Overview of the talk

- Major goal is to introduce my two main projects:
  - change point detection for time series analysis and,
  - a barycentric coding model for measures on graphs
- Unifying toolbox: the optimal transport (OT) problem and the geometry of probability distributions

Signal Processing

A Rich Tradition in Applied Mathematics

- A *signal* is just a function. Often it's a function in time, but for today we'll actually be interested in functions of space as well.
- Three tasks today: *pattern identification, analysis* of a signal and *synthesis* of a signal
- Pattern Identification: Looking for common themes in the data
- Analysis: understanding a signal in terms of some essential set of components
- Synthesis: building a new signal out of some essential set of components

Signal Processing

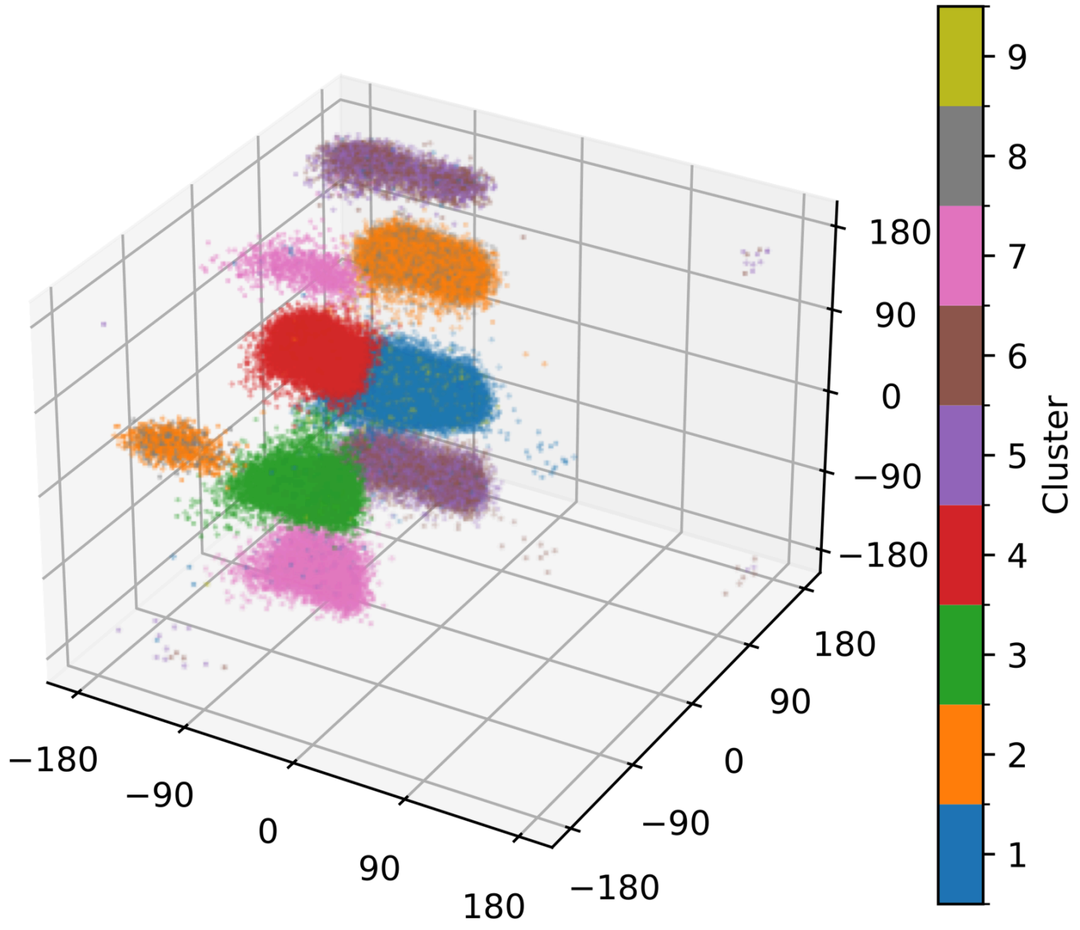
What kind of signals do we care about?

- Today, we'll restrict to a very specific kind of signal: *probability measures*.
- A probability measure is a non-negative function on some domain Ω that sums (integrates) to 1.
- The set of all probability measures on a given domain I'll denote by $\mathcal{P}(\Omega)$
- If $\Omega = \mathbb{R}^d$, we may also make reference to the space of finite variance and absolutely continuous measures $\mathcal{P}_{2,ac}(\Omega)$
- If we're willing to conceptualize the domain Ω as a physical space, we can also think of a probability measure as being a "pile of sand with unit mass", which we might want to transport around, depending on our needs

Signal Processing

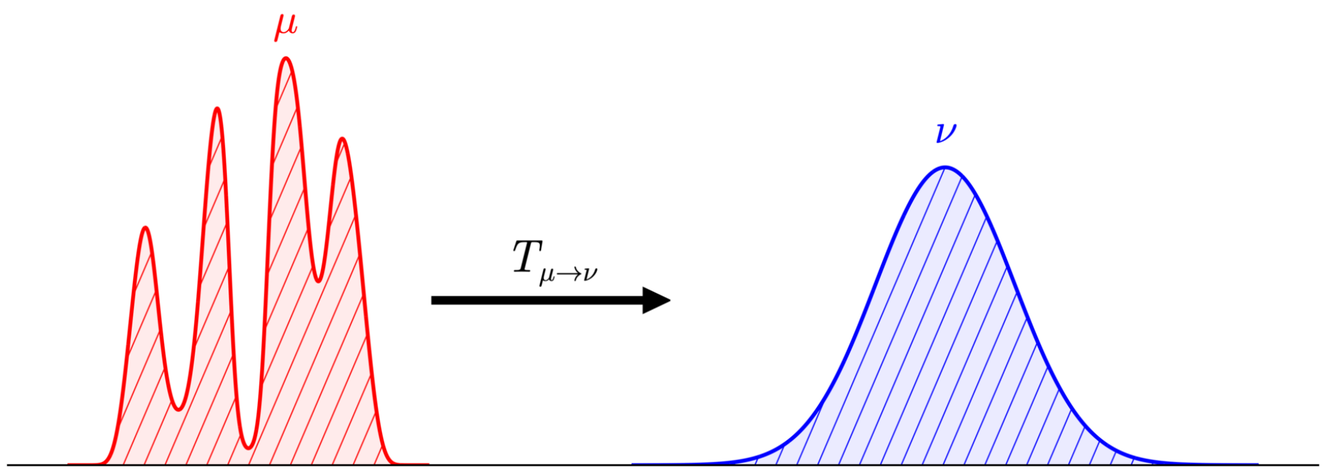
What do I mean by “spatial” signal processing?

- By “spatial” signal processing, I mean specifically any signal processing problem where the underlying geometry of the domain plays an important role
- **Major point:** because these problems are informed by the geometry of the underlying domain, we want some tool that encodes that information and makes use of it
- Enter…



Optimal Transport:

A Tool for Comparing Probability Distributions



Background on Optimal Transport

Kantorovich Problem, Wasserstein Geometry

- Let $\mu, \nu \in \mathcal{P}(\Omega)$, $c : \Omega \times \Omega \to \mathbb{R}^+$. Then the *optimal transport cost* between μ and ν is given by the minimal value of the *Kantorovich problem*:

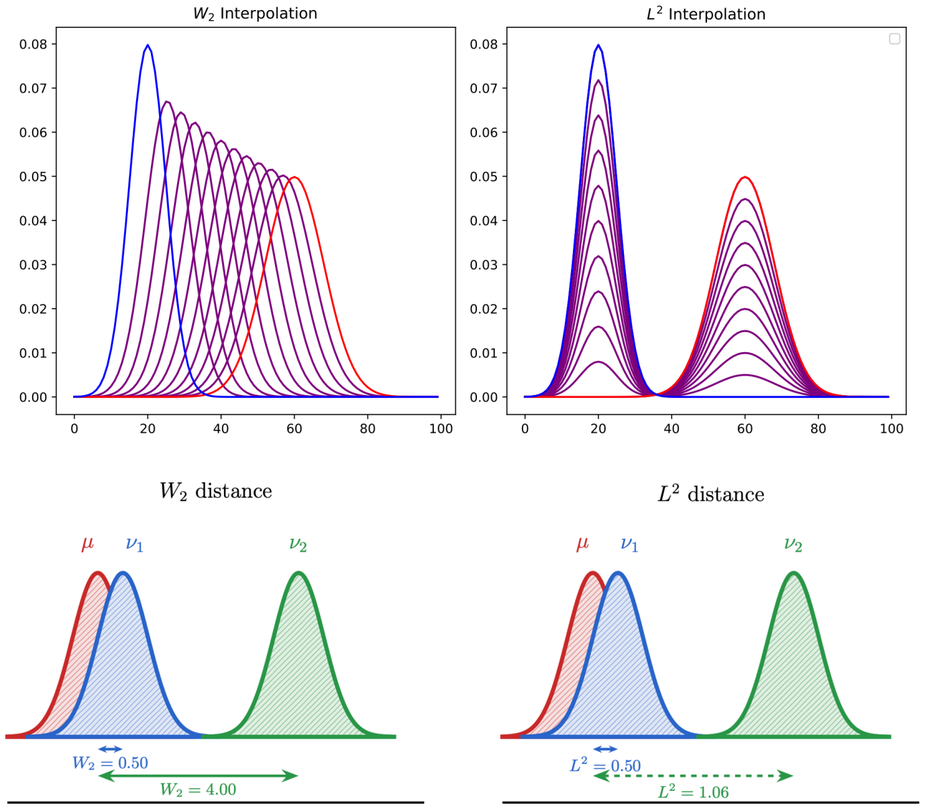
$$\mathrm{OT}(\mu, \nu) = \inf \left\{ \int_{\Omega \times \Omega} c(x, y)\, d\gamma(x, y) : \gamma \in \Pi(\mu, \nu) \right\}$$

- Remarkable fact: if $\Omega = \mathbb{R}^d$, $c(x, y) = \|x - y\|^2$, and we limit ourselves to probability measures of finite variance, then the square root of the optimal transport cost is a *metric* on $\mathcal{P}_{2,ac}(\Omega)$, denoted by W_2,

$$W_2^2(\mu, \nu) = \inf \left\{ \int_{\Omega \times \Omega} \|x - y\|^2\, d\gamma(x, y) : \gamma \in \Pi(\mu, \nu) \right\}.$$

Background on Optimal Transport

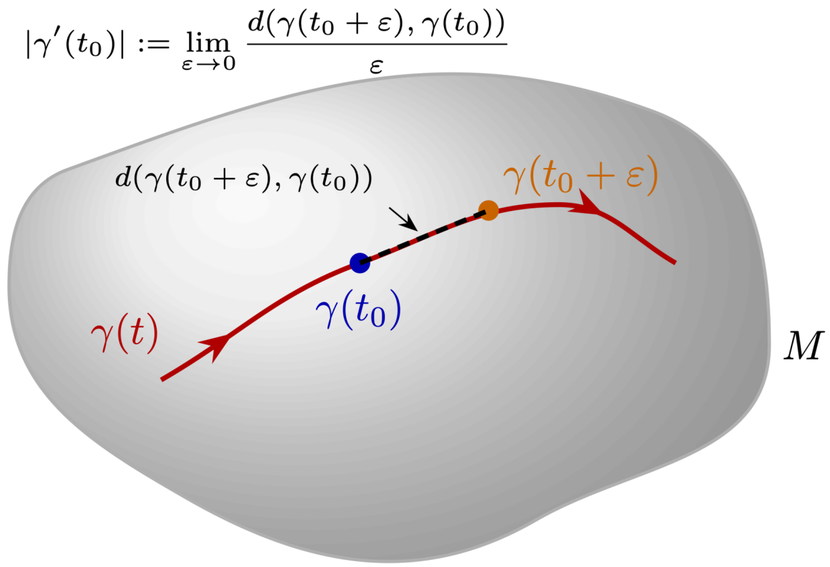
How does transport encode information about the underlying domain?



- The W_2 distance between two distributions tells us how expensive it is to move one to the other “physically”
- This contrasts with, e.g., the L^2 difference between the densities of two distributions, where mass “teleports”

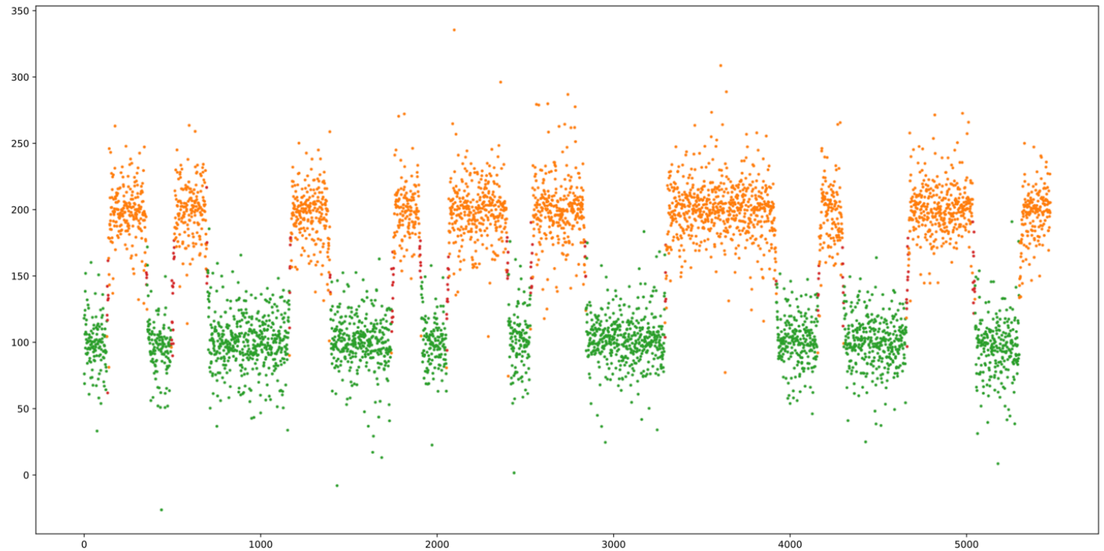
Background on Optimal Transport

Metric Derivative in the Wasserstein Geometry



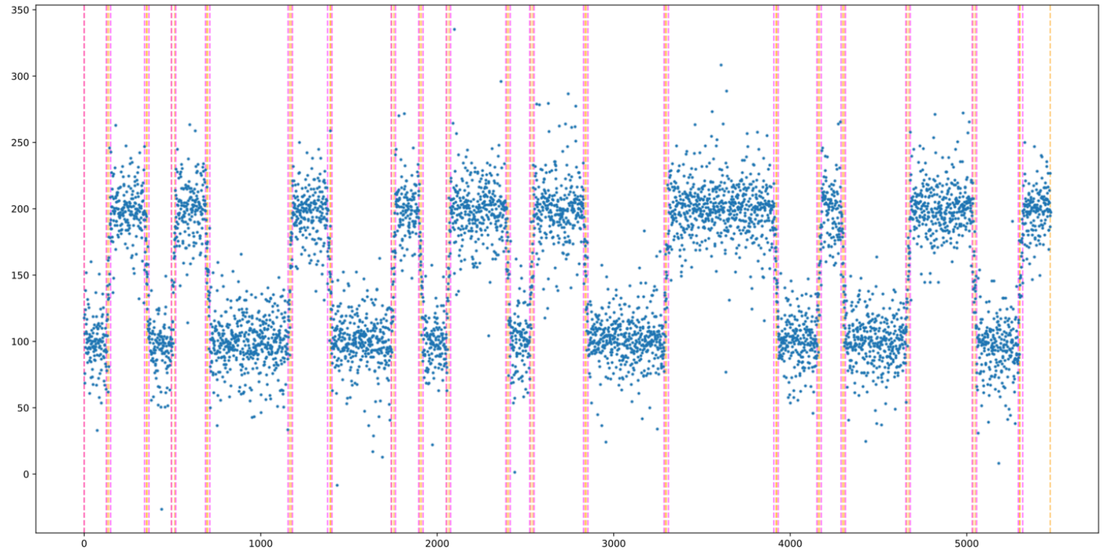
- A tool we'll use for the CPD approach is the *metric derivative*, which we will try to estimate based on data
- Intuition: difference quotient of a curve in space measuring its instantaneous speed (no direction information)
- We will use this to measure the rate at which a local ensemble of data points is evolving in the space of probability distributions!

Application: A Transport Motivated Approach to Time Series Clustering



Problem: Detecting Change Points in a Time Series

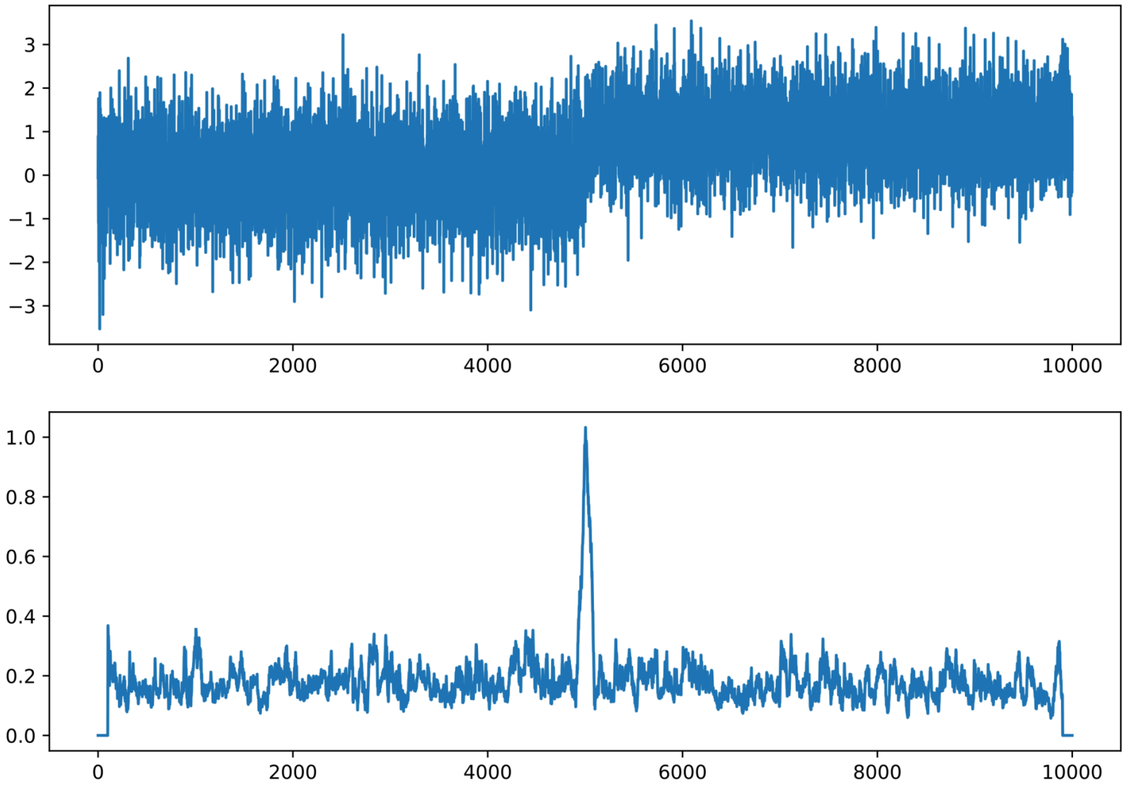
- Given a realization of a stochastic process $\{X_t\}_{t=1}^{T}$, we want to detect when significant changes to the law of the process occur
- **Main idea:** better to look at many samples simultaneously to get a better sense of the overall state of the process and robustly identify patterns and structures in the observed data



Detecting Changes with the Optimal Transport Cost

Key Idea: Changes happen when neighboring distributions differ in W_2

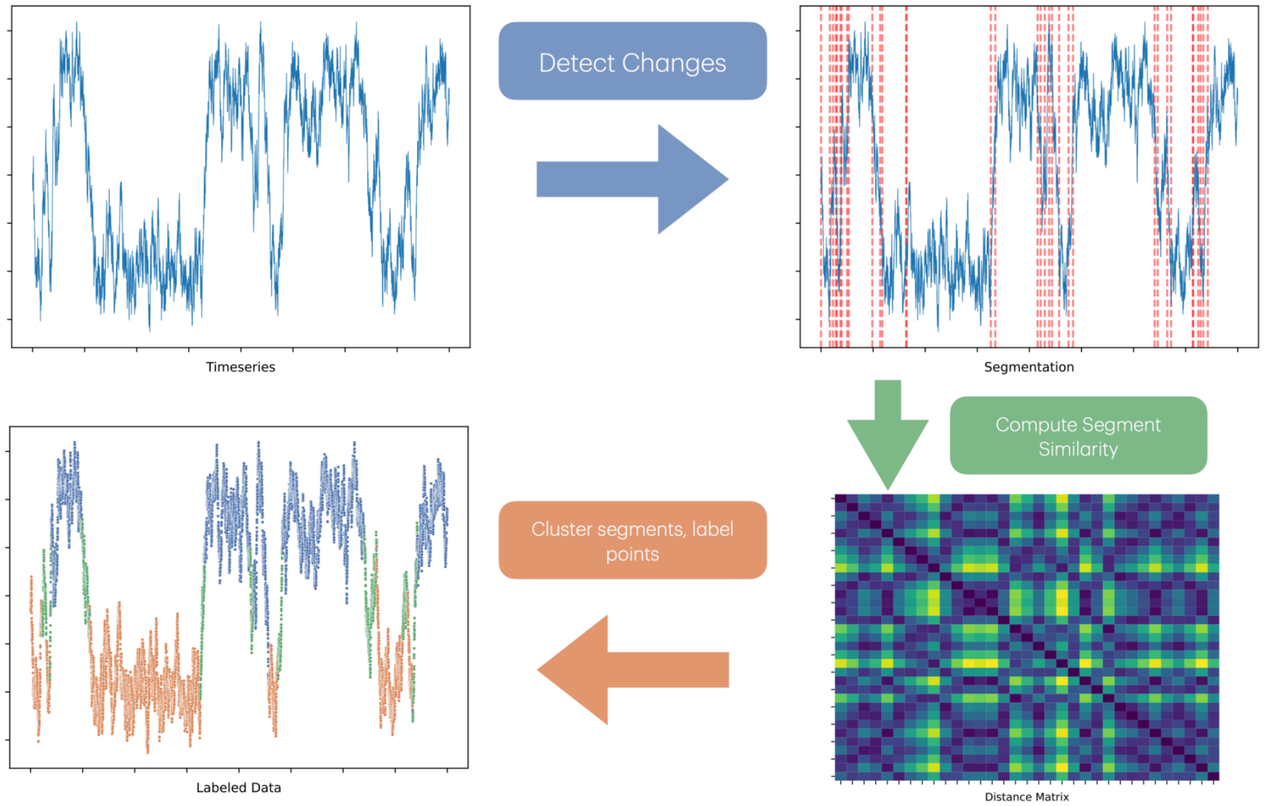
- We use a sliding window approach to look for CPs
- At each step we approximate the metric W_2 derivative between neighboring batches of w samples
- Then a quantile-based cutoff is applied to identify CP candidates, which are further culled by identifying inflection points of the metric derivative data in contiguous segments



Clustering Segments based on Change Points

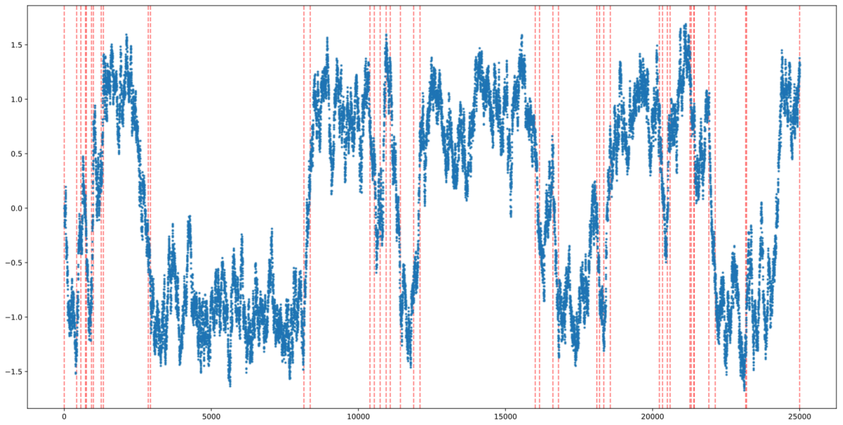
Key Idea: Quantify Segment Similarity with W_2

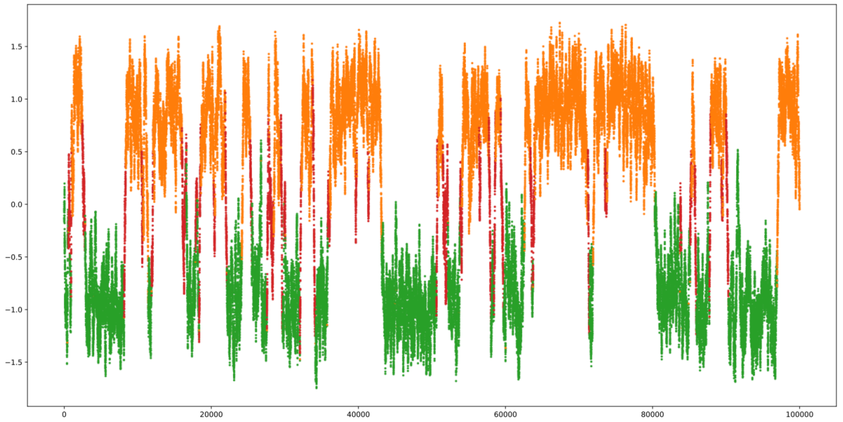
- We compute pairwise segment similarity by measuring their distance in W_2
- Then we use a “classical” clustering algorithm (ADPC) that analyzes the pairwise distance matrix
- For multidimensional data, we carry this procedure out on each component of the data and then glue the resulting labels together



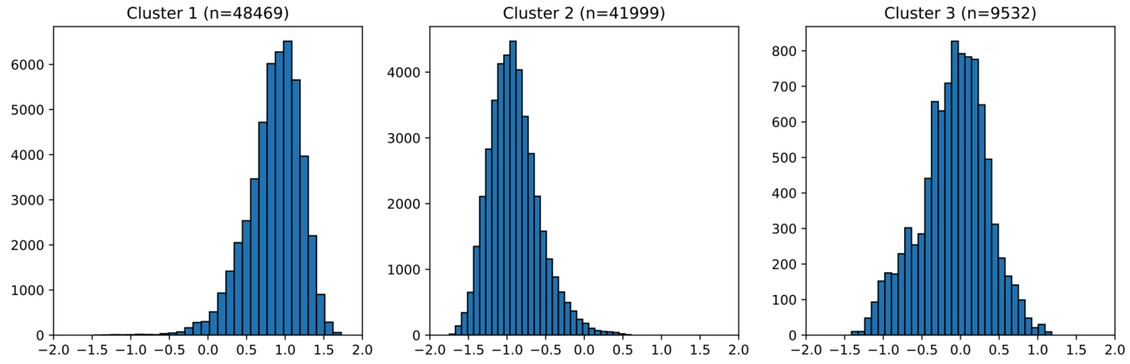
Numerical Experiment

Clustering a particle trajectory driven by Langevin dynamics



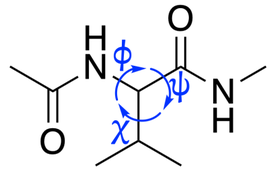


- Example: trajectory of a particle in a double-well potential
- Clustering reveals both wells and a transition state between the two meta-stable ones

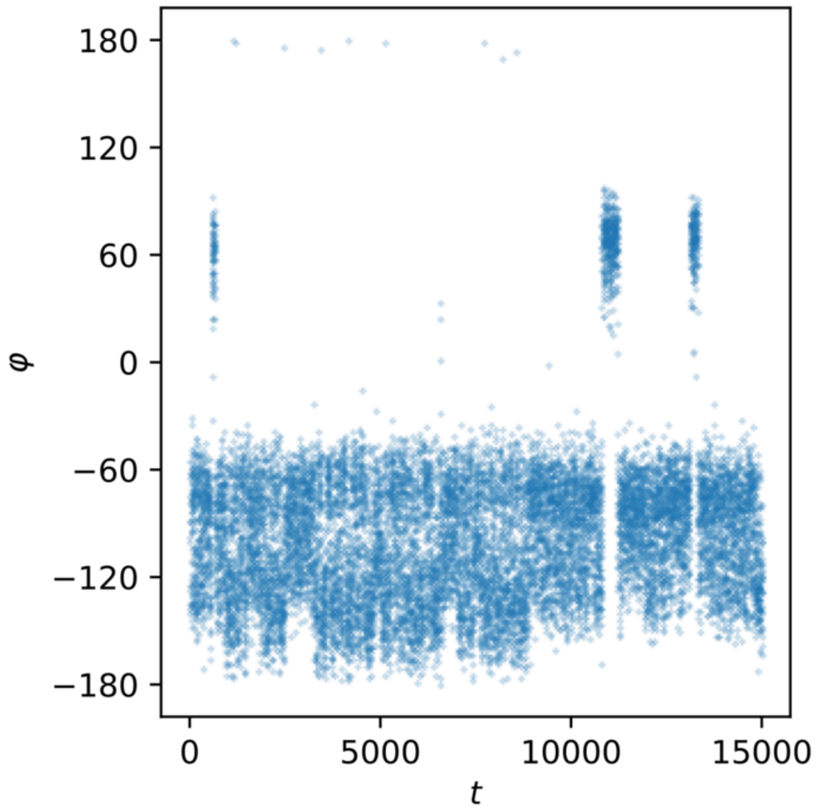


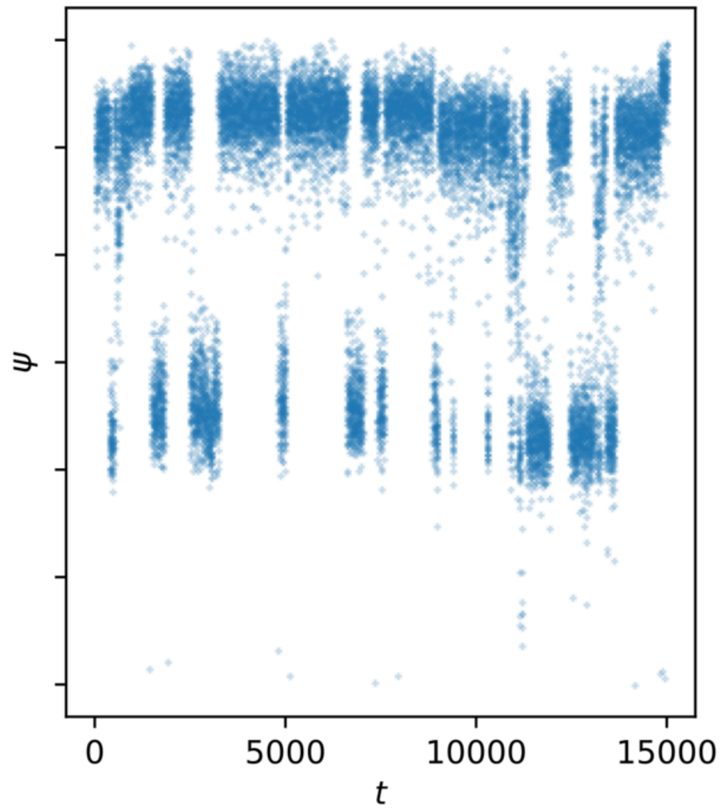
Numerical Experiment

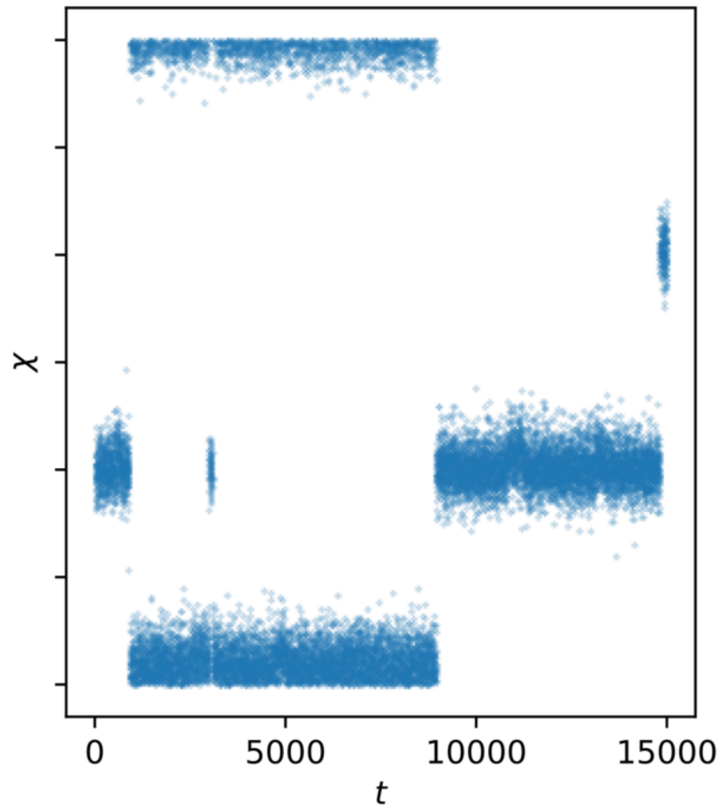
Clustering A Valine Dipeptide Trajectory



2-acetamido-*N*,3-dimethylbutanamide
Chemical Formula: $C_8H_{16}N_2O_2$

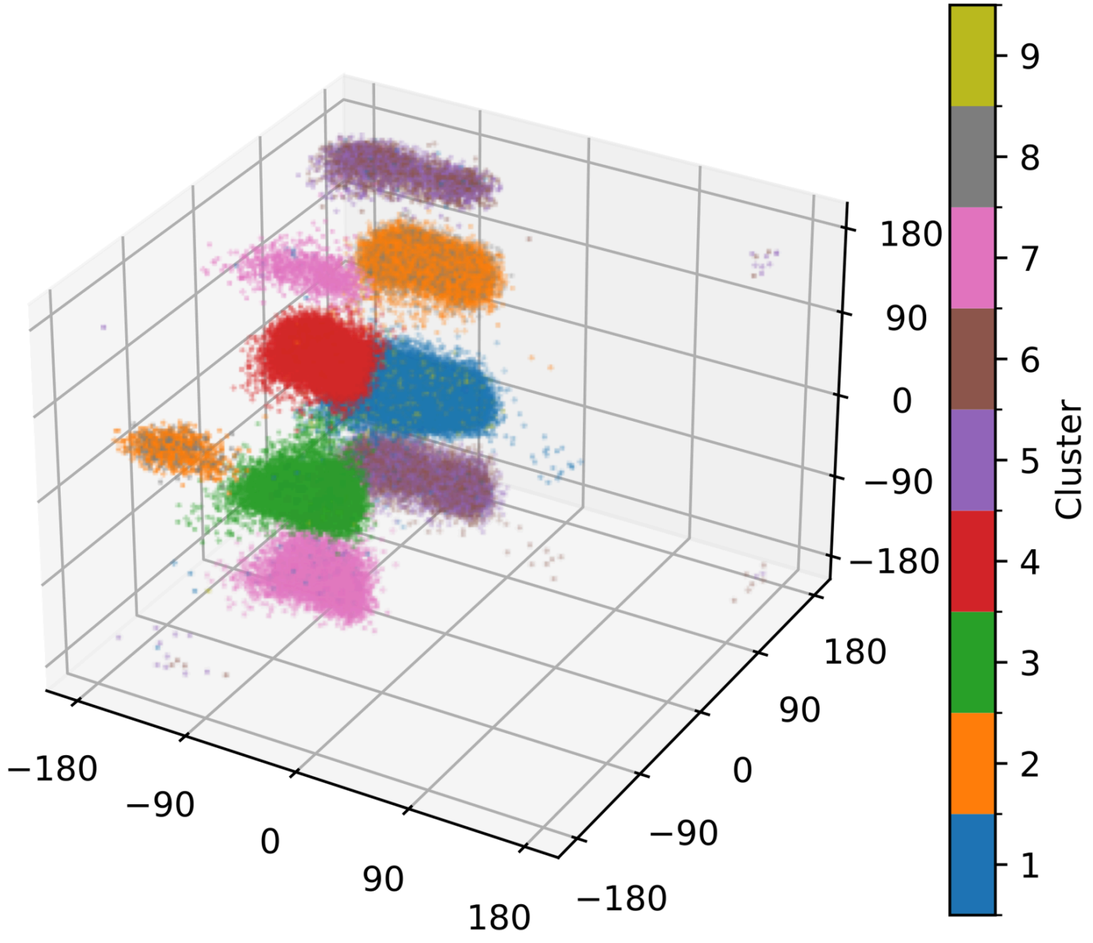


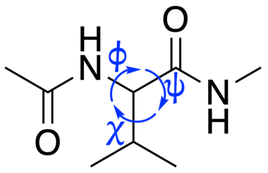




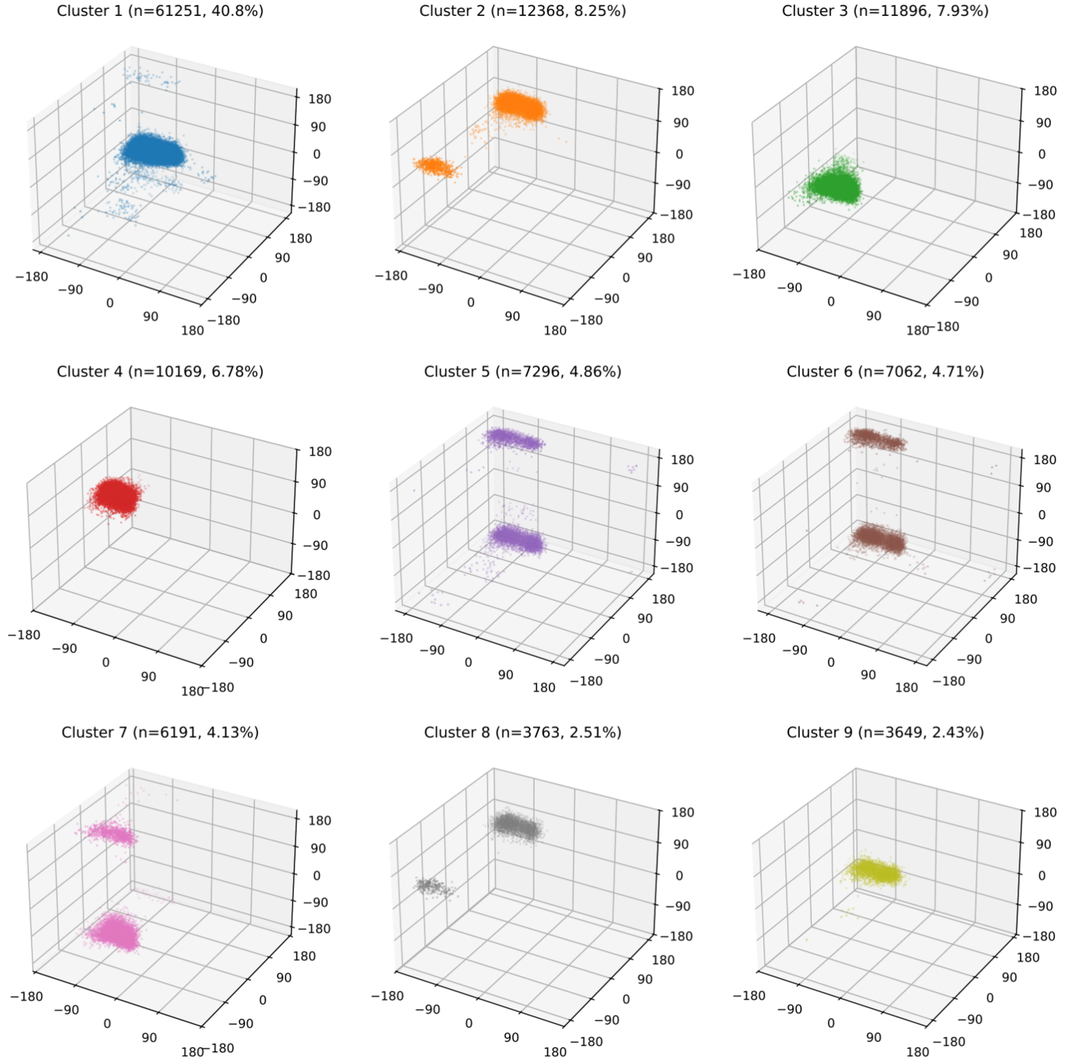
Numerical Experiment

Clustering A Valine Dipeptide Trajectory



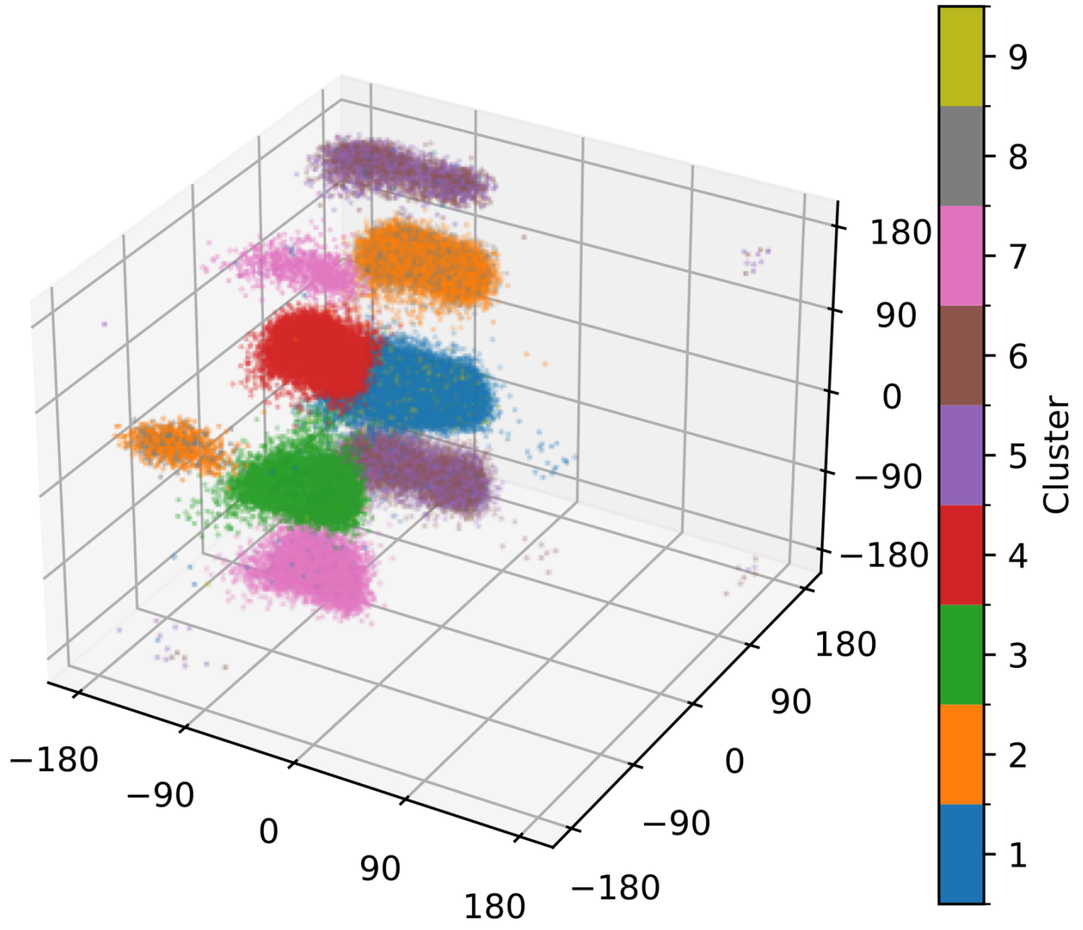


2-acetamido-*N*,3-dimethylbutanamide
Chemical Formula: $C_8H_{16}N_2O_2$



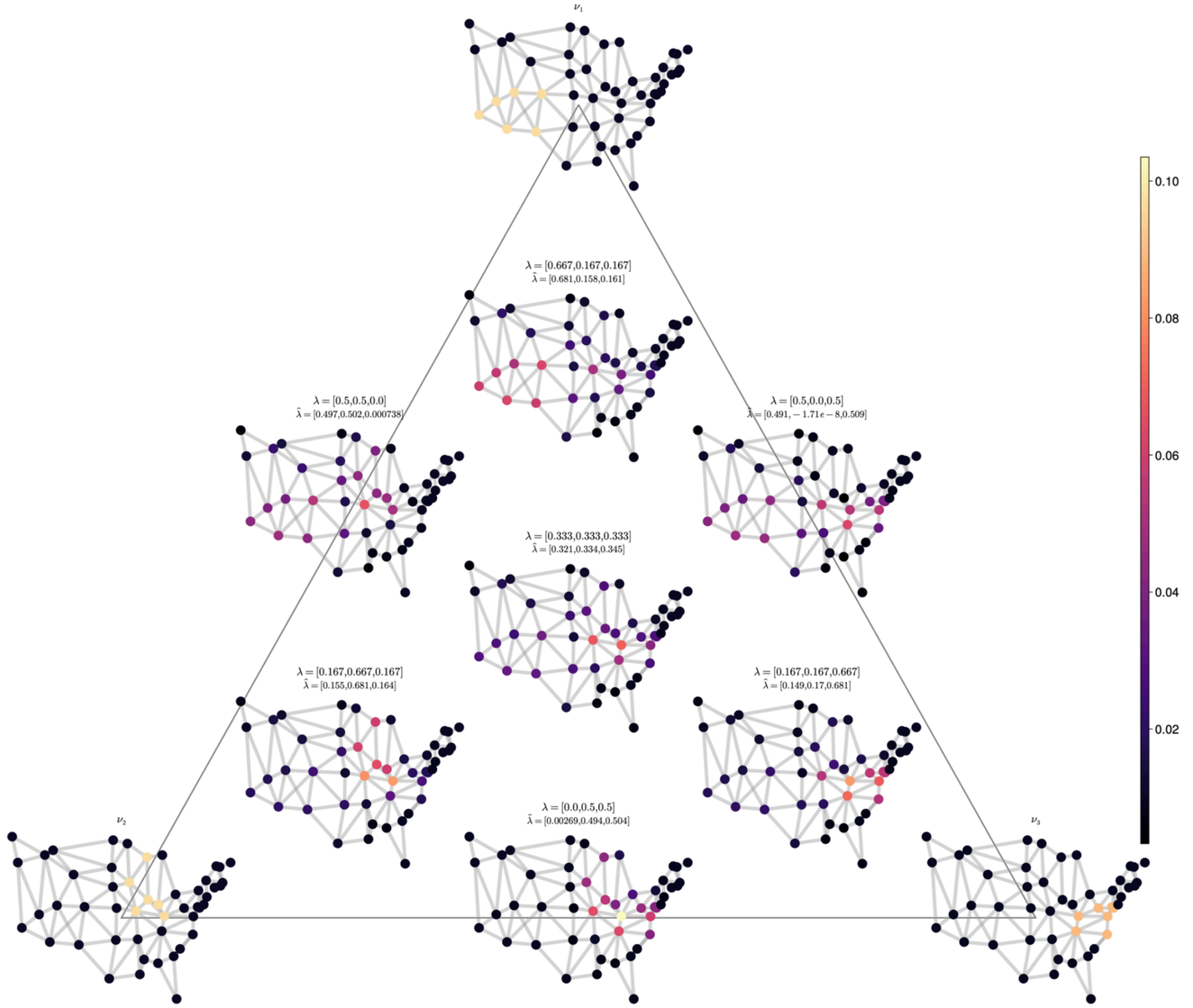
Takeaways

For Applications of Optimal Transport to CPD



- We applied optimal transport to identify from data patterns in the phase space of an observed stochastic process
- Did this by concocting an algorithm that identifies a segmentation of a given time series in a non-parametric way (i.e., no assumptions on the underlying distribution)
- We cluster the segments in W_2 space and get natural looking clusters for the trajectory in its phase space

A Barycentric Coding Model for Measures on Graphs



Structure for this part of the talk

Switching gears

1. Motivate the barycentric coding model (BCM) in general
2. Explain why transport on *graphs* requires some extra machinery
3. Show how we leverage that machinery to implement the BCM
4. Highlight some visual results from our numerical experiments
5. Conclude with musings about future directions this project might take

What the heck is a barycentric coding model?

And why is it for “measures”?

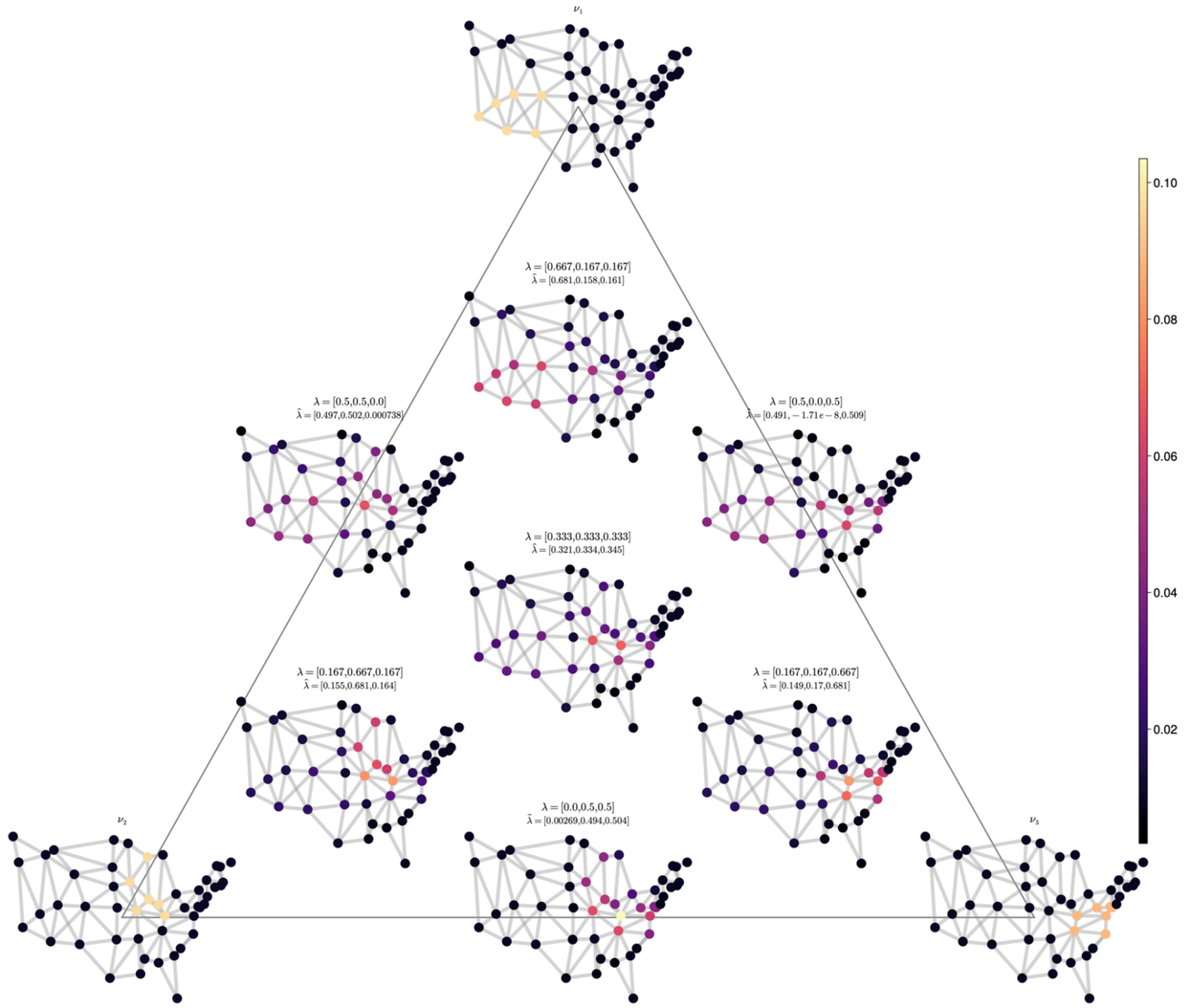
- A *barycentric coding model* is a way of using a set of essential components (“dictionary atoms”) $\mathcal{R} = \{\nu_i\}_{i=1}^p$ to *construct* and *analyze* a broad family of measures in terms of simplicial weights $\lambda \in \Delta^{p-1}$

$$\Delta^{p-1} = \{(\lambda_1, \ldots, \lambda_p) : \sum \lambda_i = 1,\ \lambda_i \geq 0\}$$

- A barycenter of *measures* is a weighted center of mass *in the transport geometry*.
- Two jobs we want to do numerically:
  - Synthesis: ID the weighted center of mass ν_λ, given $\lambda \in \Delta^{p-1}$ and $\mathcal{R} = \{\nu_i\}_{i=1}^p$
  - Analysis: ID the weights $\lambda \in \Delta^{p-1}$ for given ν_0, $\mathcal{R} = \{\nu_i\}_{i=1}^p$

Example BCM

- Vertices of the triangle correspond to dictionary atoms
- Points inside the triangle are syntheses of the measures on the vertices
- Measures are annotated with their “true” coordinates and the coordinates recovered by our analysis algorithm.



Definitions

Graphs & Markov Transition Kernels

- A *graph* is an example of a discrete metric space.
- A graph is defined by *vertices* and *edges*, and we typically write something like $X = (V, E)$, sometimes with an associated (symmetric) weight matrix $\omega : E \to \mathbb{R}^+$ (if unspecified, $\omega \equiv 1 \; \forall (x, y) \in E$)
- We will assume our graphs are simple, and can be represented by a Markov transition kernel Q with unique steady state distribution π

$$\pi(x) = \frac{\deg_\omega(x)}{\sum_{y \in V} \deg_\omega(y)}; \qquad Q(x, y) = \begin{cases} \dfrac{\omega(x, y)}{\deg_\omega(x)} & (x, y) \in E \\ 0 & \text{else.} \end{cases}$$

Definitions

Functions, vectors, probability measures on graphs

- A function φ on X is a function on its vertices and therefore a vector in $\mathbb{R}^{|V|}$
- A vector field Φ on X is a function on its edges, and therefore a vector in $\mathbb{R}^{|V|\times|V|}$
- For technical reasons, we'll consider probability measures *with respect to the steady vector* π, i.e.,

$$\mathcal{P}(X) = \left\{ \nu \in \mathbb{R}_+^{|V|} : \sum_{x \in X} \nu(x)\pi(x) = 1 \right\}$$

Barycenters, Classic Edition

Motivation: Convex combinations in $\mathbb{R}^d$

- In $\mathbb{R}^d$, given points $\{x_i\}_{i=1}^p$ and weights $\lambda = (\lambda_1, \ldots, \lambda_p) : \sum \lambda_i = 1$, it's easy to combine the points:

$$x = \sum_{i=1}^{p} \lambda_i x_i$$

- But lots of spaces lack the linear structure that enables this straightforward combination
- Idea: observe that x solves a certain optimization problem:

$$x = \arg\min \sum \frac{\lambda_i}{2} \|x - x_i\|^2$$

Barycenters Classic Edition

Motivation: Barycenters of Measures in the W_2 geometry

- To define a weighted average ν_λ of *measures* $\mathcal{R} = \{\nu_i\}_{i=1}^p \subset \mathcal{P}_{2,ac}(\Omega)$ for given weights $\lambda \in \Delta^{p-1}$, we look for minimizers of sums of squared distances, solving:

$$\inf \left\{ J[\nu] = \sum_{i=1}^{p} \frac{\lambda_i}{2} W_2^2(\nu, \nu_i) \right\}$$

- We'll sometimes refer to J as the *variance functional*. It's a major theorem that in the Euclidean setting, minimizers of J exist and are unique under appropriate technical conditions, see Agueh & Carlier (2011) for details.

Agueh, Carlier. "Barycenters in the Wasserstein space." *SIAM Journal on Mathematical Analysis* 43.2 (2011): 904–924

Why is this even a problem?

Degeneracy of the Classical Transport Geometry when $|\Omega| < \infty$

- Consider the case $X_2 = \{a, b\}$, then every measure $\nu \in \mathcal{P}(X_2)$ looks like $\nu = \nu_\alpha = (1 - \alpha,\ 1 + \alpha),\ \alpha \in [-1, 1]$
- Exercise to the reader: use a little linear algebra to characterize every coupling for a fixed pair of measures $\nu_\alpha, \nu_\beta \in \mathcal{P}(X_2)$ and solve the K.P. explicitly
- Conclude that $(\mathcal{P}(X_2), W_2) \simeq \left([0, 1], \sqrt{|x - y|}\right)$
- Further conclude that if $\nu(t) : [0, 1] \to \mathcal{P}(X_2)$ is a geodesic curve then it is constant! So you can't even define a barycenter of two measures on a two point graph. Sad fact of life.

More Background on Optimal Transport

Benamou-Brenier Formula

- Cool transport fact: W_2 can be interpreted as arising from a certain Riemannian structure on an "infinite dimensional manifold", via the *Benamou-Brenier formula* (see Benamou & Brenier (2000) for details)

$$W_2^2(\rho_0, \rho_1) = \inf \left\{ \mathcal{A}(\rho(x,t), V(x,t)) = \int_0^1 \|V(t,x)\|^2_{L^2(\rho(t,x))}\, dt \right\}$$

- The infimum is taken over pairs $\rho : [0,1] \times \mathbb{R}^d \to \mathbb{R}$, $V : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$ which satisfy the *continuity equation with temporal boundary conditions:*

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho V) = 0; \qquad \rho(0,\cdot) = \rho_0,\ \rho(1,\cdot) = \rho_1$$

Benamou, Jean-David, and Yann Brenier. "A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem." *Numerische Mathematik* 84.3 (2000): 375–393.

Continuity Equations on Graphs

The admissible mean

- Notice: need to make sense of the product ρV in the CE

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho V) = 0$$

- An *admissible mean* is a function $\theta : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$; it must be concave, lower semi-continuous, 1-homogeneous, symmetric, and C^2 on $(0, \infty)^2$.
- Examples:

$$\theta_{\log}(s,t) = \frac{s-t}{\log(s) - \log(t)}; \qquad \theta_{geo}(s,t) = \sqrt{st}$$

- For $\rho \in \mathcal{P}(X)$, define $\theta(\rho)(x,y) = \theta(\rho(x), \rho(y))$; now θ turns functions into vector fields

Continuity Equations on Graphs, Cont’d

Finally: we write down a continuity equation

- We now have the language to express a *discrete continuity equation*

$$\mathcal{CE}(\rho_A, \rho_B) := \left\{ (\rho, m) : \frac{\partial \rho}{\partial t} + \operatorname{div}_X(\theta(\rho) \odot m) = 0,\ \rho(0,\cdot) = \rho_A,\ \rho(1,\cdot) = \rho_B \right\}$$

$$\mathcal{A}(\rho, m) := \frac{1}{2} \int_0^1 \sum_{i,j \in X} \frac{m(t,i,j)^2}{\theta(\rho(t,i), \rho(t,j))}\, Q(i,j)\pi(i)\, dt.$$

$$\mathcal{W}(\rho_A, \rho_B) := \inf \left\{ \sqrt{\mathcal{A}(\rho, m)} : (\rho, m) \in \mathcal{CE}(\rho_A, \rho_B) \right\}.$$

Maas, Jan. “Gradient flows of the entropy for finite Markov chains.” *Journal of Functional Analysis* 261.8 (2011): 2250–2292.

The $\mathcal{W}$ Geometry

Getting back towards barycenters

- Recall that the variance functional is expressed directly in terms of a metric!

$$J[\nu] = \sum_{i=1}^{p} \frac{\lambda_i}{2} \mathcal{W}^2(\nu_i, \nu)$$

- We want to minimize it, and to do that we'll use one of the oldest tricks in the book: gradient descent
- Main idea: look for the direction in which J is decreasing most rapidly, and take a "small step" in that direction
- Two problems now: 1) how do I know what direction to walk in and 2), how do I walk in it?

The $\mathcal{W}$ Geometry, Cont'd

Riemannian Geometry Objects: Tangent Space & Metric Tensor

- We'll need to understand two things about the Riemannian geometry of $(\mathcal{P}(X), \mathcal{W})$: the metric tensor, and the exponential mapping
- First, we need to say what the tangent space is, and how we can compute the inner product of two vectors at a point. This turns out to be relatively easy!

$$T_\nu \mathcal{P}(X) = \left\{ \nabla\varphi \in \mathbb{R}^{|V|\times|V|} : \varphi \in \mathbb{R}^{|V|} \right\},$$

$$\langle \nabla\varphi, \nabla\psi \rangle_\nu = \frac{1}{2} \sum_{x,y\in V} \theta(\nu(x), \nu(y))\, \nabla\varphi(x,y)\, \nabla\psi(x,y).$$

The $\mathcal{W}$ Geometry, Cont'd

Riemannian Geometry Objects: Exponential Mapping and Inverse

- Now on any generic Riemannian manifold, one can consider the *exponential mapping* and its inverse, which relates the local geometry of the manifold to its tangent space

$$\exp_\nu : U \subset T_\nu \mathcal{P}(X) \to \mathcal{P}(X),$$

- If $\gamma_{\nu,m} : [0,1] \to \mathcal{P}(X)$ is a geodesic curve with $\gamma(0) = \nu,\ \gamma'(0) = m$, then

$$\exp_\nu(m) = \gamma_{\nu,m}(1),$$

and

$$\exp_\nu^{-1}(\gamma_{\nu,m}(1)) = m.$$

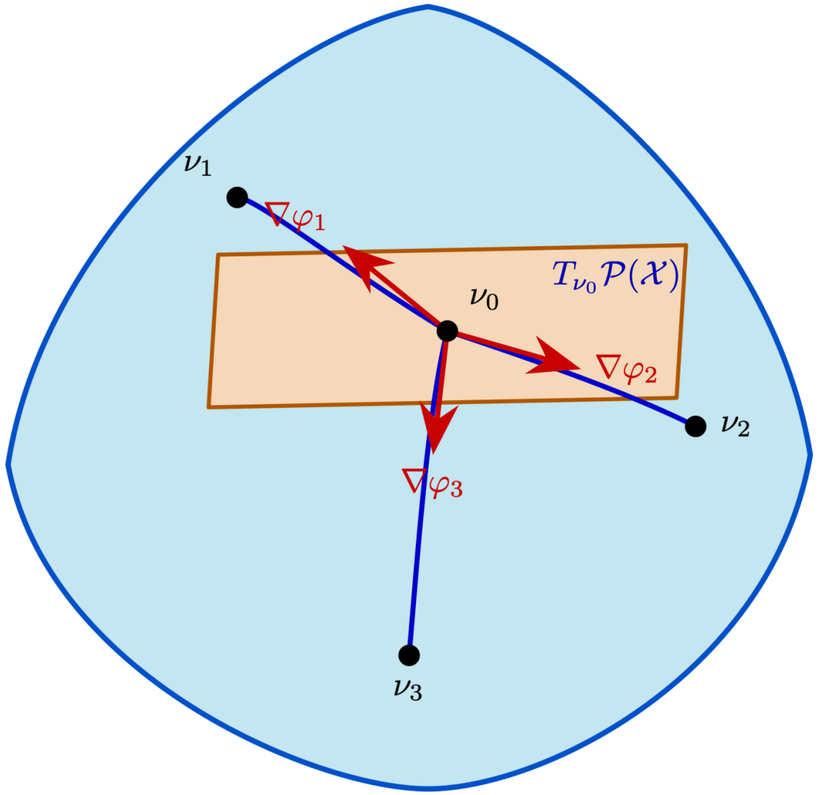
Synthesis Plan: Intrinsic Gradient Descent

When in Non-Euclidean spaces, do as the Non-Euclidean Optimizers Do

- Classic way to optimize a function: gradient descent

$$x_{k+1} = x_k - t_k \nabla F(x_k)$$

- Compute the gradient, walk in the direction of steepest descent, eventually find a minimum where $\nabla F(x_\infty) = 0$
- We'll do something similar here, but to keep ourselves on the set $\mathcal{P}(X)$, we will use the *intrinsic gradient* associated to a certain Riemannian structure



Analysis Plan: Inverting The Exponential Map

Take it to the Tangent Space

- In order to analyze measures with respect to a family of references $\mathcal{R}$, we leverage a theorem due to Werenski et al.

Theorem (Werenski et al. (2022)). *Let $\{\nu_i\}_{i=1}^p$ be a collection of suitably regular probability measures on Ω. Then $\nu_0 \in \mathrm{Bary}(\{\nu_i\}_{i=1}^p)$ if and only if*

$$\min_{\lambda \in \Delta^{p-1}} \lambda^T A \lambda = 0, \tag{1}$$

where, letting T_i be the optimal transport map from ν_0 to ν_i, $A \in \mathbb{R}^{p \times p}$ is defined by

$$A_{ij} = \langle T_i - \mathrm{Id}, T_j - \mathrm{Id} \rangle_{L^2(\nu_0)}. \tag{2}$$

Furthermore, if the minimum value of $\lambda^T A \lambda$ is 0 and λ^* is an optimal argument, then $\nu_0 = \nu_{\lambda^*}$.

Werenski, Matthew, et al. “Measure estimation in the barycentric coding model.” *International Conference on Machine Learning*. PMLR, 2022.

Ambrosio, Luigi, Nicola Gigli, and Giuseppe Savaré. *Gradient flows: in metric spaces and in the space of probability measures.* Basel: Birkhäuser Basel, 2005.

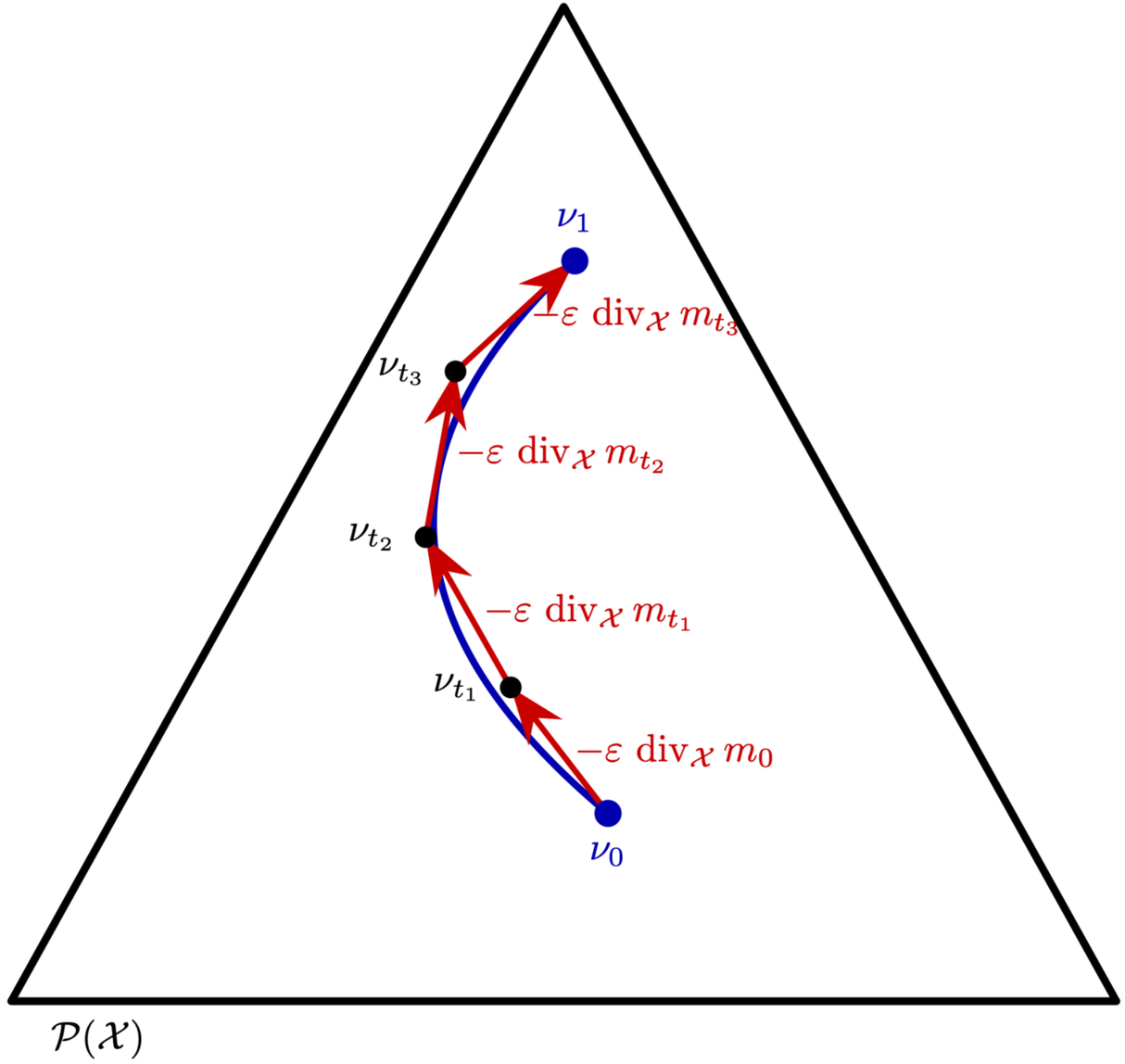
Putting it All Together

Summary of Needs

- So in short we need:
  - To be able to *differentiate the variance functional* J, in order to identify the direction to walk in when synthesizing a measure
  - To be able to *approximate the exponential mapping* at a measure $\nu \in \mathcal{P}(X)$ in order to “walk forward” towards the final synthesis measure
  - To be able to *approximate the inverse of the exponential mapping* to lift manifold points into the tangent space, and form the quadratic program which we will solve to analyze measures

Key Idea:

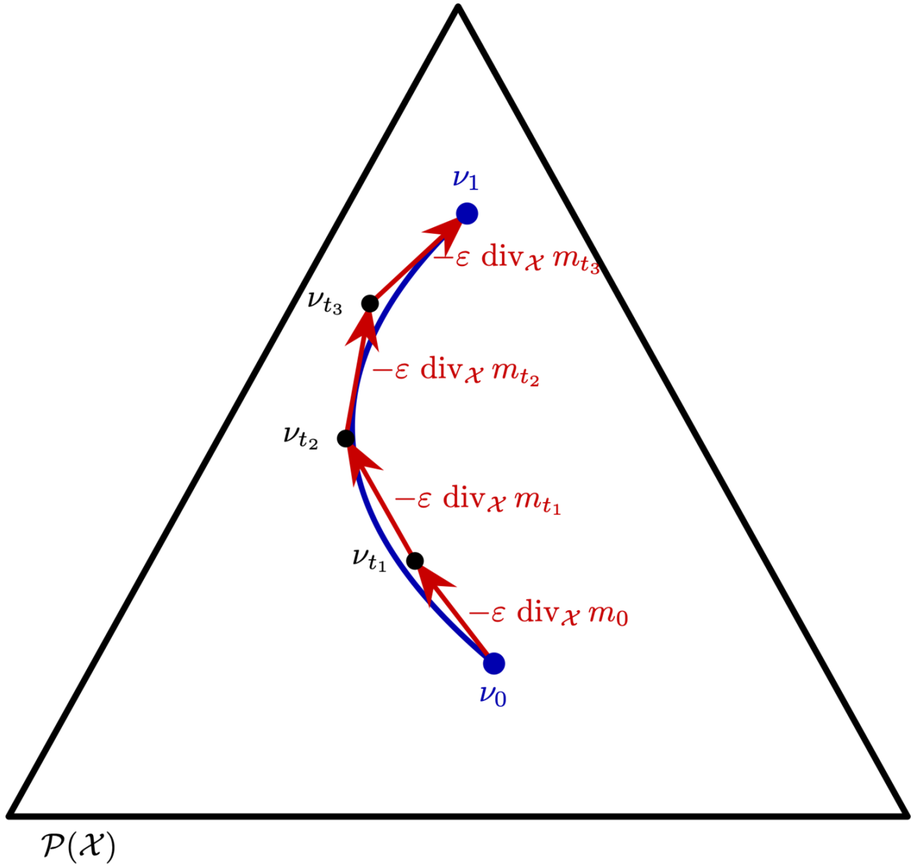
Approximating Geodesic Curves in the $\mathcal{W}$ geometry



Approximating Geodesic Curves via Chambolle-Pock

One Neat Trick

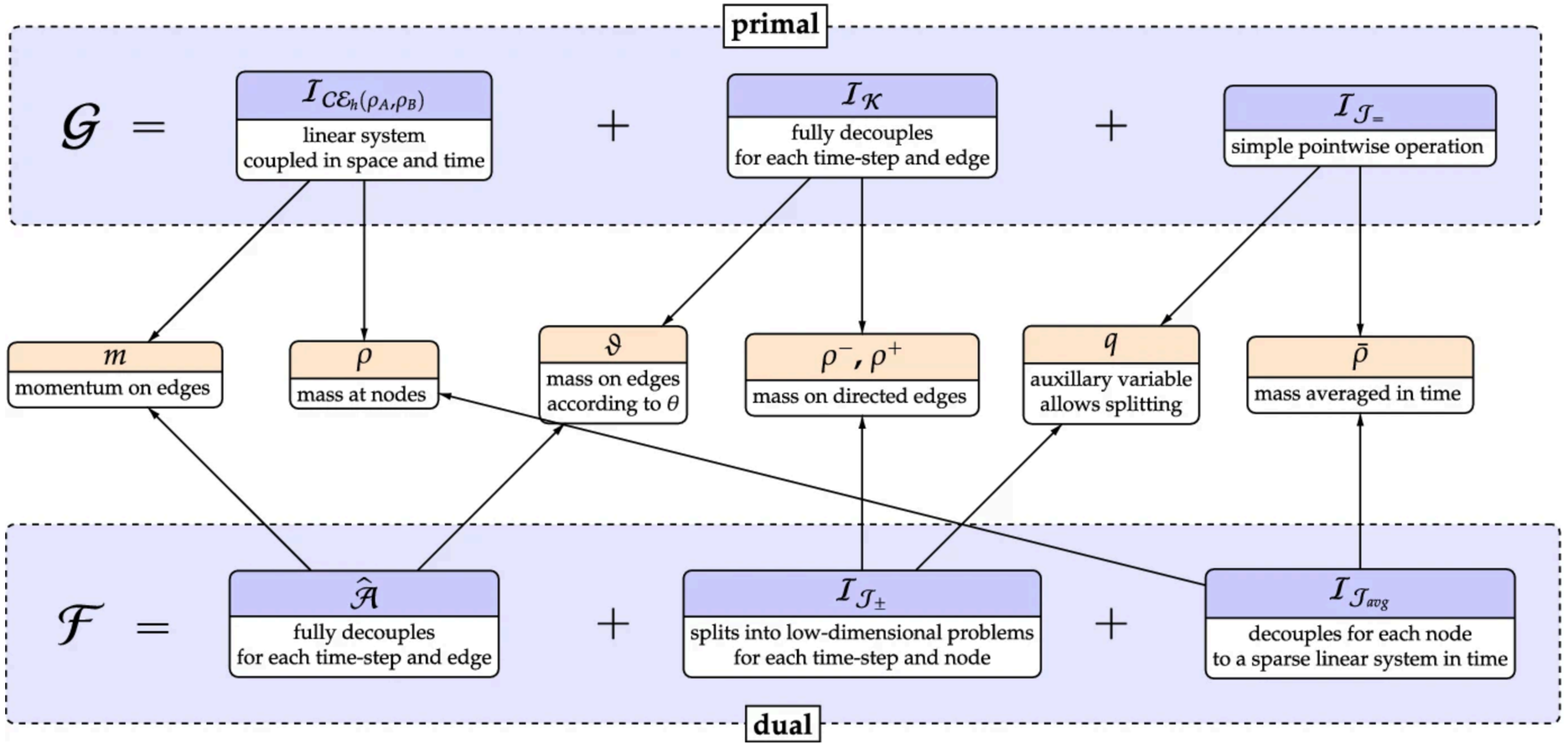
- Erbar et al. (2020) proposed a method for approximating the numerical value of $\mathcal{W}$ by applying a non-smooth convex optimization algorithm
- At a high level: use a weak formulation of the CE, introduce slack variables to decouple variables, minimize the action via Chambolle-Pock



Erbar, Matthias, et al. “Computation of optimal transport on discrete metric measure spaces.” *Numerische Mathematik* 144.1 (2020): 157–200.

Approximating Geodesic Curves via Chambolle-Pock

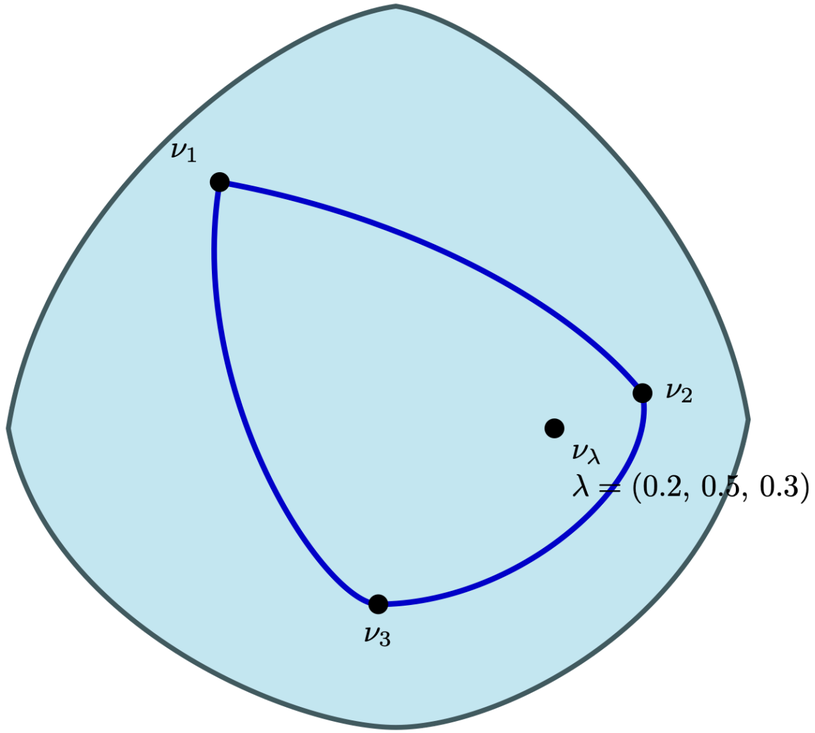
It’s simple, really.



Erbar, Matthias, et al. “Computation of optimal transport on discrete metric measure spaces.” *Numerische Mathematik* 144.1 (2020): 157–200.

Synthesizing Barycenters of Measures

Key idea: choose the update direction using Riemannian calculus

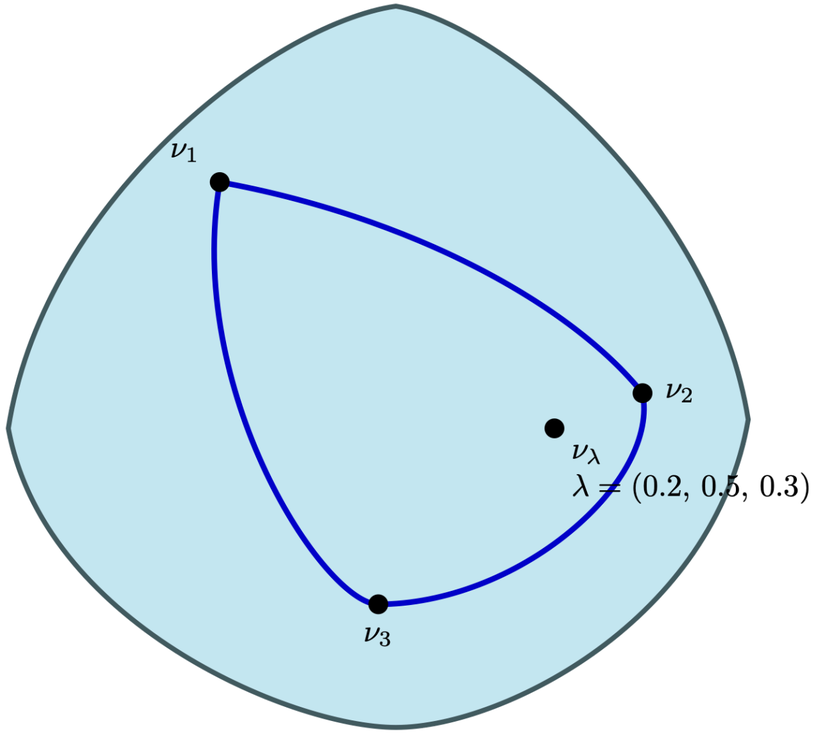


- The direction we need to walk in is the direction in the tangent space which most rapidly decreases J
- We can use generic results from Riemannian geometry to differentiate J

$$\nabla_{\mathcal{W}} \left\{ \sum_{i=1}^{p} \frac{\lambda_i}{2} \mathcal{W}^2(\nu_i, \nu) \right\} = -\sum_{i=1}^{p} \lambda_i \exp_{\nu}^{-1}(\nu_i)$$

Synthesizing Barycenters of Measures

Key idea: evolve forward via the continuity equation



- In order to update our descent iterates we need to be able to "walk in the direction of steepest descent", i.e. evaluate the exponential map at least near the current iterate
- Our update step will therefore be (for $0 < t_k \ll 1$)

$$\nu_{k+1} = \exp_{\nu_k}\left(-t_k\left(\sum_{i=1}^{p}\lambda_i \exp_{\nu_k}^{-1}(\nu_i)\right)\right)$$

- Observation: geodesic curves must obey the discretized continuity equation
- Idea: use a heuristic update scheme based on the continuity equation near the "launch point" (current iterate)

Synthesizing Barycenters of Measures

Update iterates with the continuity equation

- If (ρ, m) is a pair satisfying the continuity equation, then we have

$$\lim_{\varepsilon \to 0} \frac{\rho(t+\varepsilon) - \rho(t)}{\varepsilon} = -\operatorname{div}_X(m(t))$$

- So take $\varepsilon > 0$ to be small, set $t = 0$, and re-arrange to obtain

$$\rho(\varepsilon) \approx \rho(0) - \varepsilon \operatorname{div}_X m(0)$$

- Idea: this update is an approximation of the exponential map near the probability measure $\rho(0)$

Synthesizing Barycenters of Measures

An algorithm for synthesis of a barycenter

Algorithm 1 Synthesis of Barycenters via Intrinsic Gradient Descent in the $\mathcal{W}$ Metric

Input: Markov chain (Q, π), reference measures $\{\nu_i\}_{i=1}^{p}$, weights $\lambda \in \Delta^{p-1}$, convergence criterion

Output: Barycenter $\nu = \mathrm{Bary}(\{\nu_i\}_{i=1}^{p}, \lambda)$

$\nu \leftarrow \nu_1$ ▷ Initialize with a measure which is trivially in $\mathrm{Bary}(\{\nu_i\}_{i=1}^{p})$

while not converged **do**

$m \leftarrow 0 \in \mathbb{R}^{|V| \times |V|}$

for $r = 1, \ldots, p$ **do**

$m \leftarrow m + \lambda_r m_{\nu,\nu_r}(0)$ ▷ Compute the initial momentum of the geodesic $[\nu, \nu_r]$ and sum

end for

$\nu \leftarrow \nu + \varepsilon \, \mathrm{div}_{\mathcal{X}} \, m$ ▷ Update via continuity equation heuristic

end while

return ν

Analyzing Barycenters

An algorithm for analysis of a measure relative to references

Algorithm 2 Analysis of Barycenters via Quadratic Programming

Input: Markov chain (Q, π), target measure ν, reference measures $\{\nu_i\}_{i=1}^p$

Output: Coordinates λ^* such that $\nu = \text{Bary}(\{\nu_i\}_{i=1}^p, \lambda^*)$

$A \leftarrow 0 \in \mathbb{R}^{p \times p}$

for $i = 1, \ldots, p$ **do**

$\nabla\varphi_i \leftarrow m_{\nu,\nu_i}(0)$ ▷ Compute the initial momentum of the geodesic $[\nu, \nu_i]$

end for

for $i, j = 1, \ldots, p$ **do**

$A_{i,j} \leftarrow \frac{1}{2} \sum_{x,y \in \mathcal{X}} \theta(\nu(x), \nu(y)) \nabla\varphi_i(x,y) \nabla\varphi_j(x,y)$ ▷ Form the Gram matrix

end for

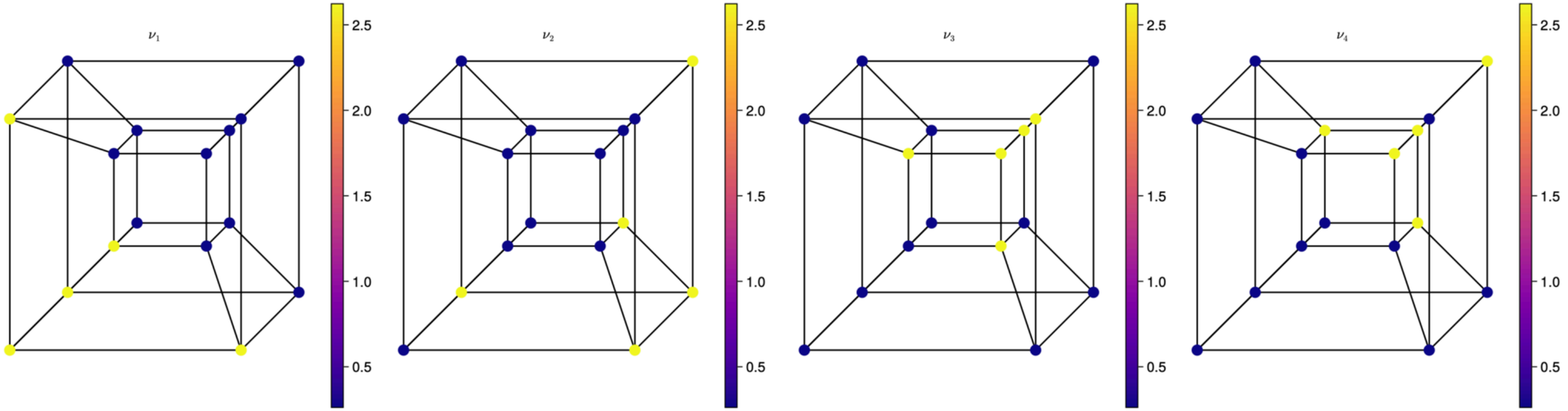
$\lambda^* \leftarrow \text{argmin} \left\{ \lambda^T A \lambda : \lambda \in \Delta^{p-1} \right\}$ ▷ Solve via CVX, etc.

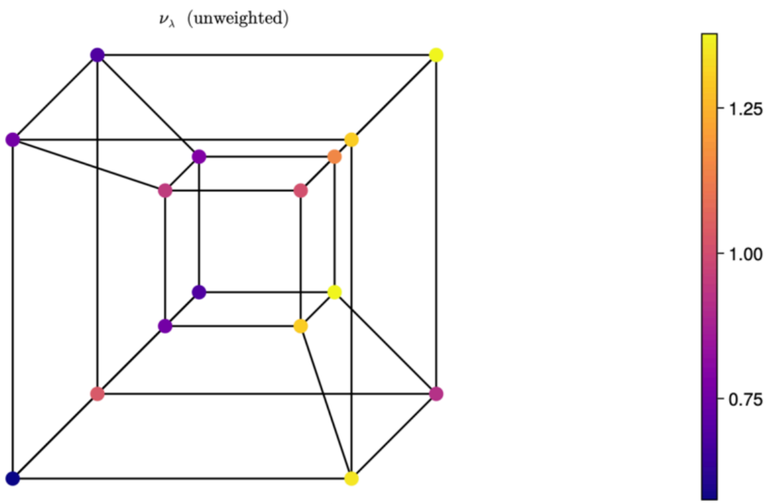
return λ^*

Hyperparameters of the Algorithms

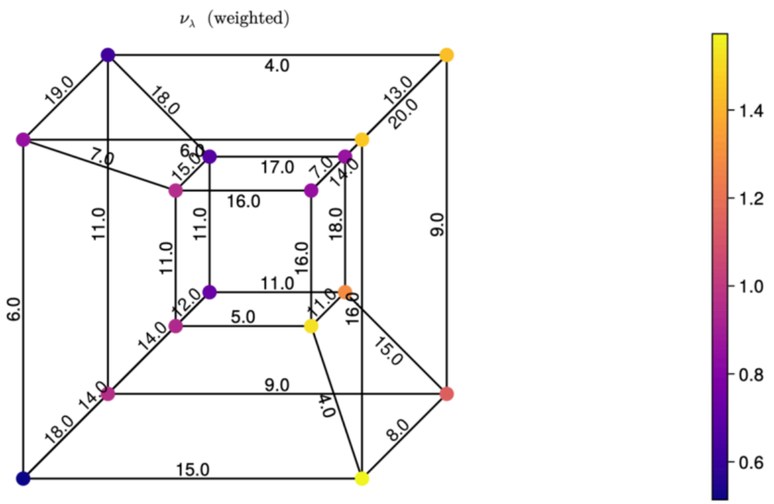
- Our algorithms involve 4 hyperparameters:
  - δ_g — convergence threshold for Chambolle-Pock
  - N — number of steps per geodesic
  - δ_b — convergence threshold for synthesis gradient descent
  - ε — step size for gradient descent
- Each geodesic is expensive to compute, with complexity scaling quadratically in both N and $|V|$

Numerical Experiments





$\lambda = [0.25, 0.25, 0.25, 0.25]$
$\hat{\lambda} = [0.25, 0.25, 0.25, 0.25]$



$\lambda = [0.25, 0.25, 0.25, 0.25]$
$\hat{\lambda} = [0.248, 0.25, 0.252, 0.249]$

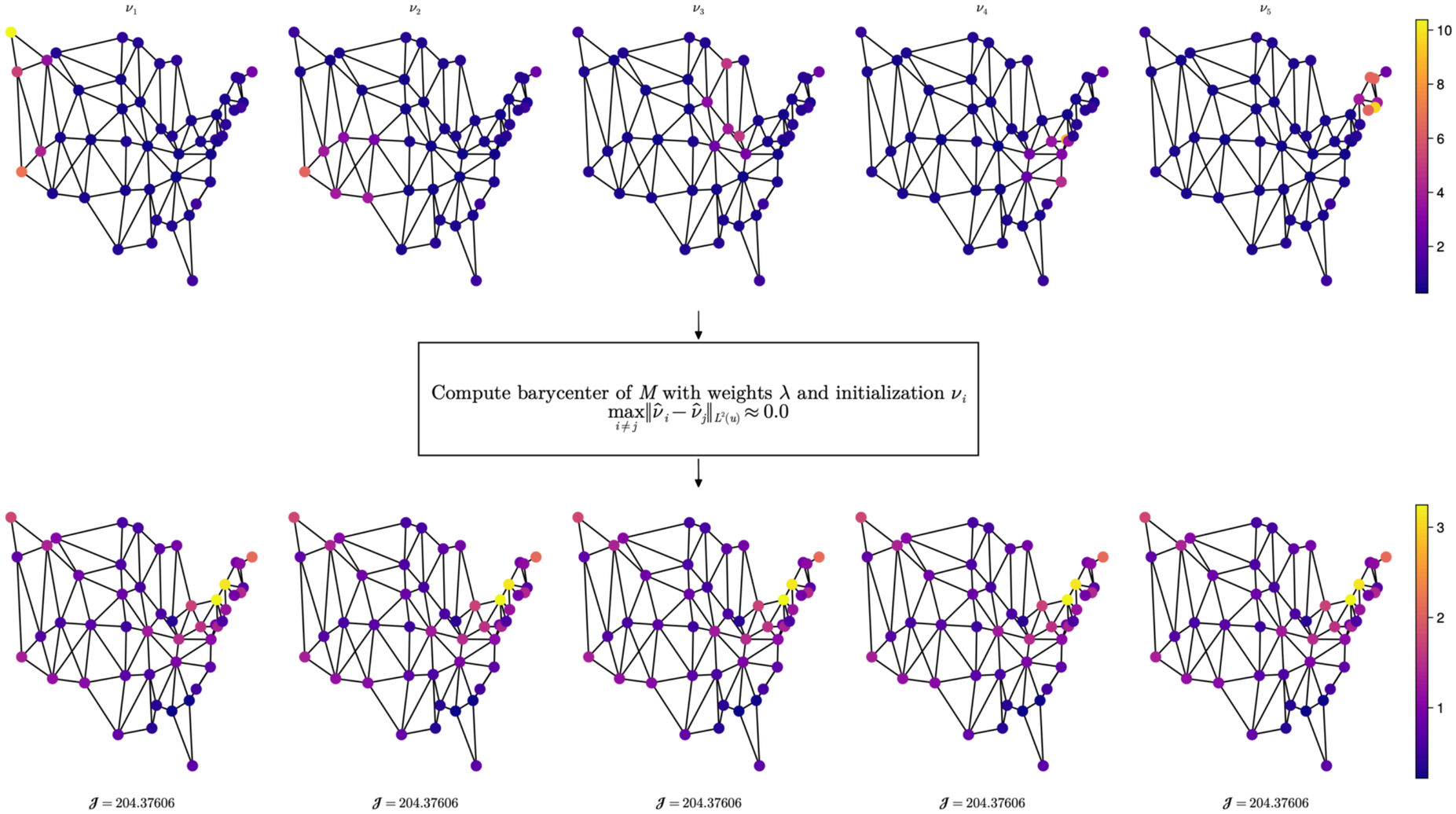
Two Important Questions

Unicity & Consistency of the Implementation

- For a fixed family of references $\mathcal{R} = \{\nu_i\}_{i=1}^{p}$, there are two important properties a BCM should have
  1. Barycenters should be uniquely specified by their coordinates λ
  2. Coordinates recovered via analysis (Alg. 2) of a synthesized measure should be consistent with the coordinates specified when synthesizing; i.e., if ν_λ is the output of Alg. 1 for inputs $(\lambda, \mathcal{R})$, and $\hat{\lambda}$ is the output of Alg. 2 for inputs $(\nu_\lambda, \mathcal{R})$, then we should have that $\lambda \approx \hat{\lambda}$
- In numerical simulations, we find both these properties to hold

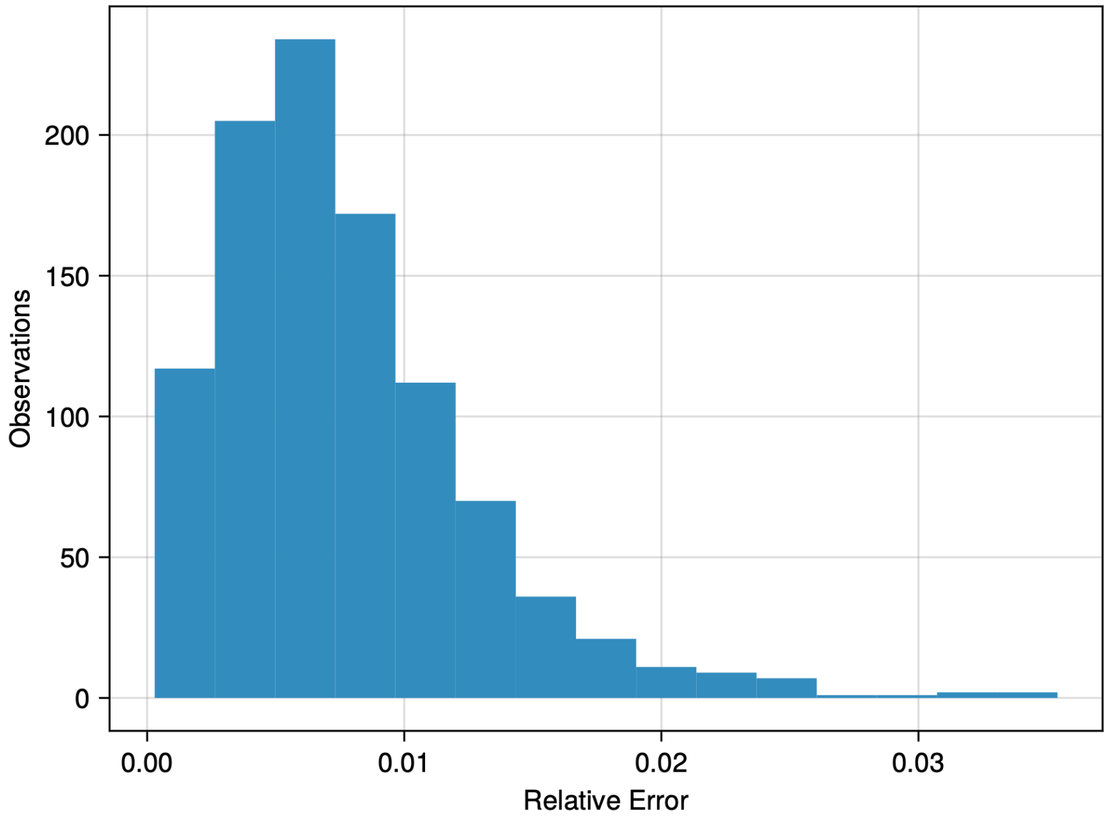
Descent Initialization

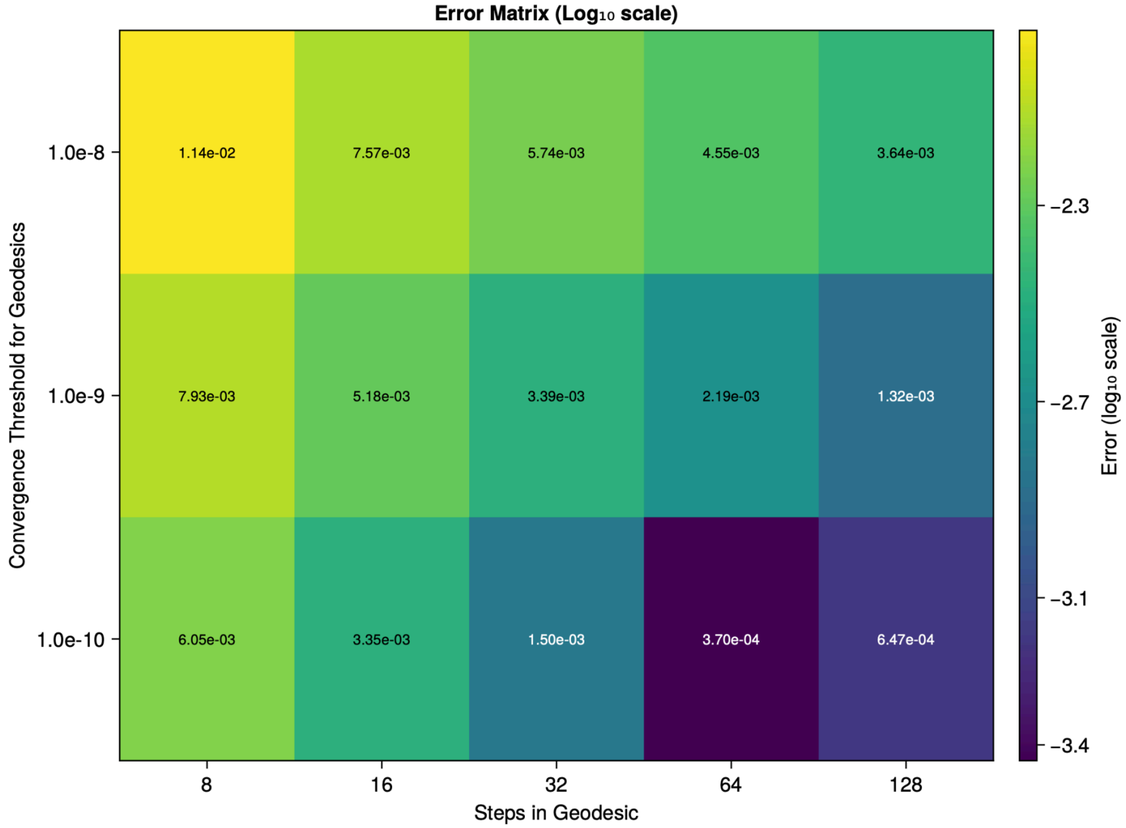
Changing the initialization used for descent doesn’t seem to impact outputs



Coordinate Recovery & Hyperparameters

Our synthesis and analysis algorithms are consistent with each other





A “Static” Approach to Barycenters of Measures

Entropic Regularization

- Popular tool in computational transport: entropic regularization
- Main thrust: regularize the Kantorovich problem with an entropy term and apply Sinkhorn’s algorithm to optimize
- How it works: fix a regularization parameter $\varepsilon_{reg} > 0$, encode the graph with a cost matrix C and solve the regularized K.P.

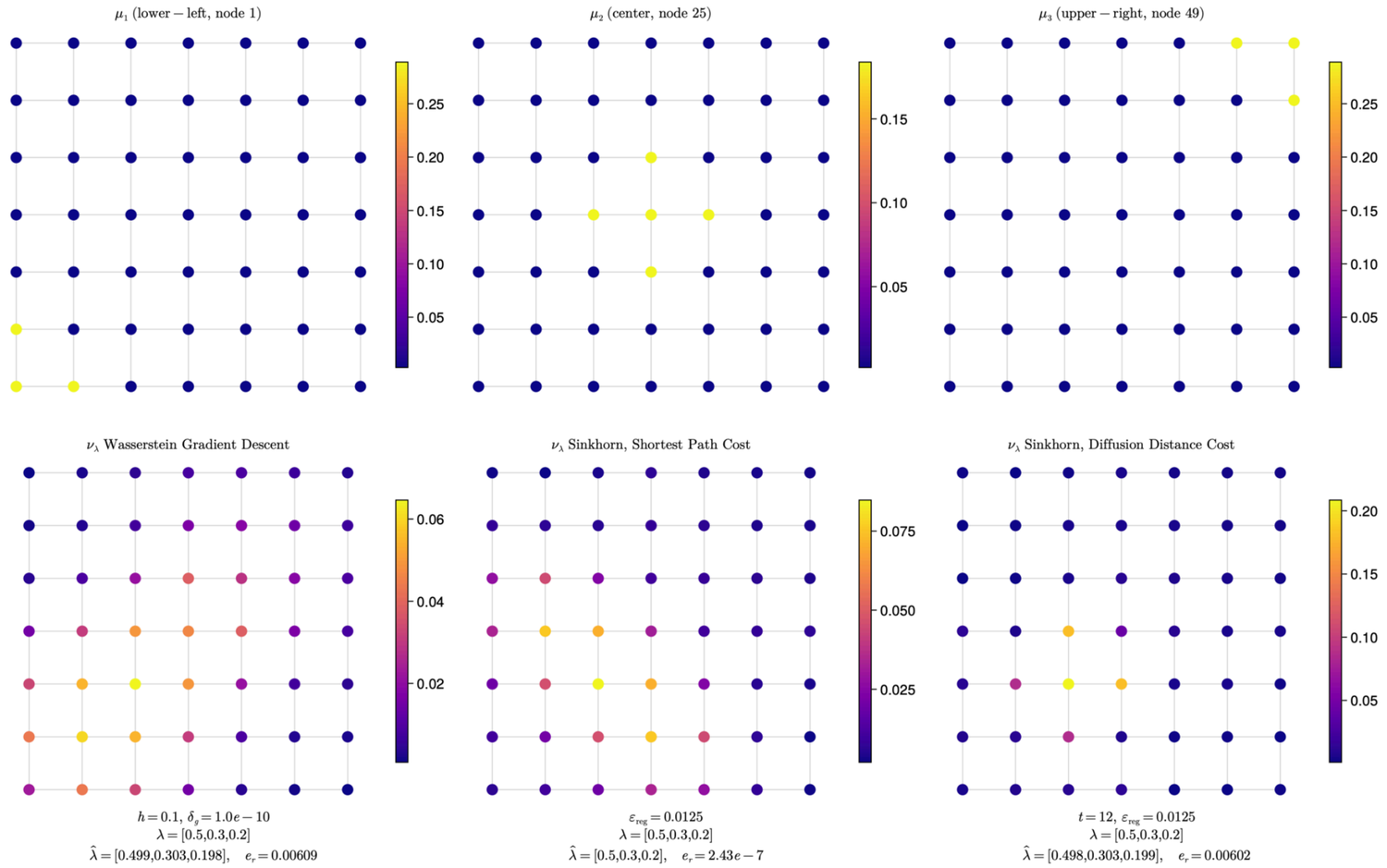
$$\mathrm{OT}_{reg}(\mu, \nu) = \inf \left\{ \langle \gamma, C \rangle + \varepsilon_{reg} H(\gamma) : \gamma \in \Pi(\mu, \nu) \right\}$$

$$H(\gamma) \equiv - \sum_{x,y \in X} \gamma(x, y) \left(\log \gamma(x, y) - 1 \right)$$

Comparison with a “Static” Approach To Computing Barycenters

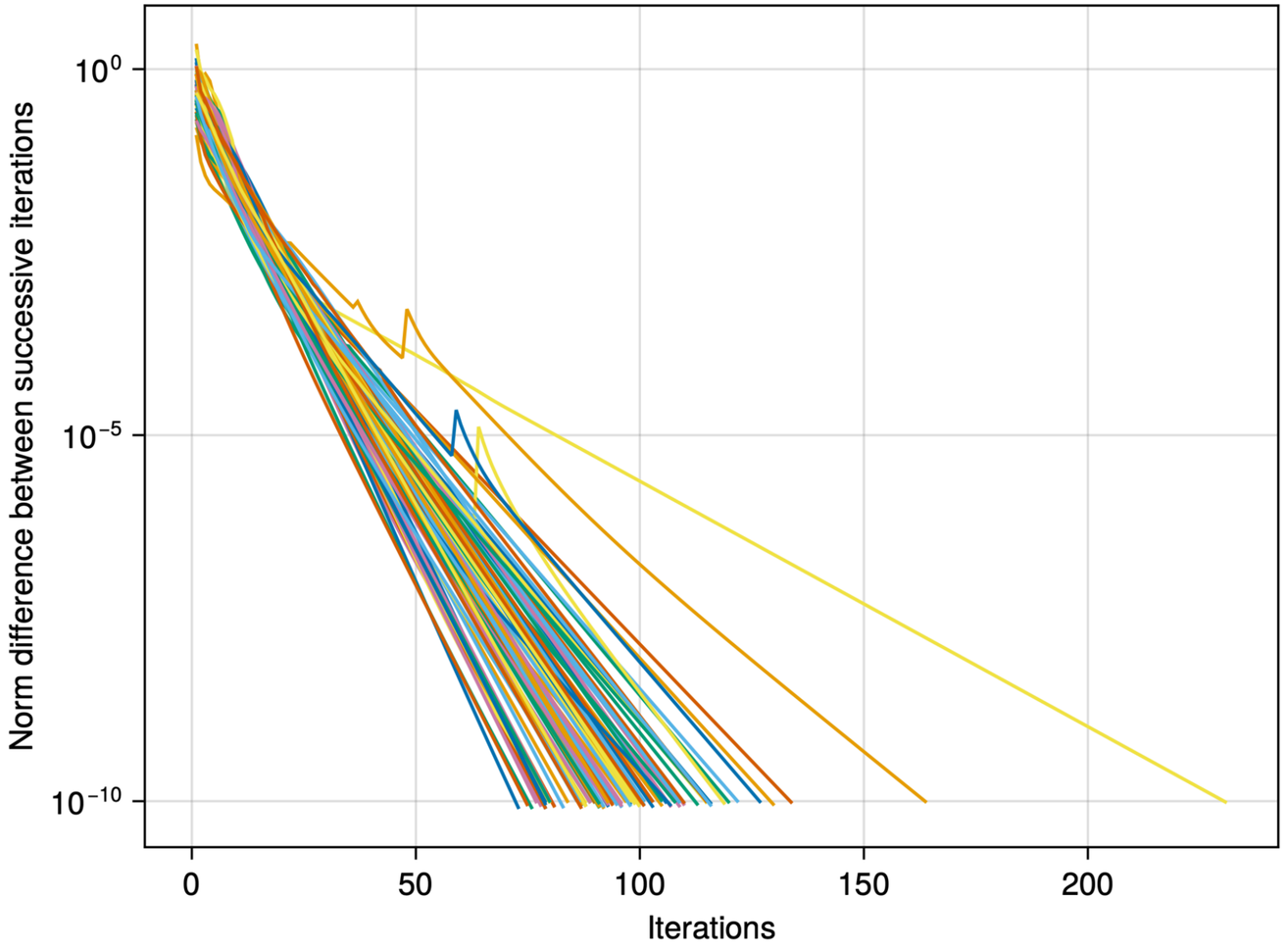
Entropic Transport

- Upsides: very fast, can have much better recovery, though not always
- Downsides: distance matrices don’t capture all connectivity information, and entropy introduces a bias to the output



Empirical convergence

- We can track the norm of the gradient over the course of descent, and we see that it is in line with what we would expect from “standard” gradient descent methods
- Possible advancement: use momentum information to accelerate descent? Area of active interest in the Riemannian optimization community



Looking Ahead

Future directions and ideas to explore

- **Key takeaway:** we can use the Riemannian geometry, our understanding of the continuity equation, and some non-smooth convex optimization algorithms to synthesize and analyze barycenters of measures on graphs
- Building out the theory:
  - Make rigorous guarantees about convergence of the descent iterates. The literature in general Riemannian optimization tells us we need more information about the injectivity radius of $(\mathcal{P}(X), \mathcal{W})$ and some estimate on the eigenvalues of the Riemannian Hessian to identify a region of guaranteed convergence for the synthesis algorithm. The latter could potentially be estimated numerically on a graph-by-graph basis.
  - Locality issues: there are graph-connectivity conditions that guarantee the support of the geodesic bridging two measures can be contained to a set $A \subset V$ — could this kind of a result be extended to the case of a general barycenter?
  - Study what happens on the boundary — it can happen that a geodesic connecting two points on the interior of the simplex touches the boundary, but can the conditions under which this occurs be directly characterized, and does it have serious implications for our numerical implementation?

References

- My papers:
  - G., Huang, and Murphy. “Wasserstein-based identification of metastable states in time series data via change point detection and segment clustering.” *arXiv preprint* *arXiv:2603.01989* (2026).
  - G., Murphy. “Static and Dynamic Approaches to Computing Barycenters of Probability Measures on Graphs.” *arXiv preprint* *arXiv:2603.26940* (2026).
- My code:
  - https://github.com/dcgentile/GraphTransportation.jl
  - https://github.com/dcgentile/mfpy
- Select references:
  - Benamou, Jean-David, and Yann Brenier. “A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem.” *Numerische Mathematik* 84.3 (2000): 375–393.
  - Agueh, Martial, and Guillaume Carlier. “Barycenters in the Wasserstein space.” *SIAM Journal on Mathematical Analysis* 43.2 (2011): 904–924.
  - Bonneel, Nicolas, Gabriel Peyré, and Marco Cuturi. “Wasserstein barycentric coordinates: histogram regression using optimal transport.” *ACM Trans. Graph.* 35.4 (2016): 71-1.
  - Werenski, Matthew, et al. “Measure estimation in the barycentric coding model.” *International Conference on Machine Learning*. PMLR, 2022.
  - Maas, Jan. “Gradient flows of the entropy for finite Markov chains.” *Journal of Functional Analysis* 261.8 (2011): 2250–2292.
  - Erbar, Matthias, et al. “Computation of optimal transport on discrete metric measure spaces.” *Numerische Mathematik* 144.1 (2020): 157–200.
  - Gangbo, Wilfrid, Wuchen Li, and Chenchen Mou. “Geodesics of minimal length in the set of probability measures on graphs.” *ESAIM: Control, Optimisation and Calculus of Variations* 25 (2019): 78.
  - Afsari, Bijan, Roberto Tron, and René Vidal. “On the convergence of gradient descent for finding the Riemannian center of mass.” *SIAM Journal on Control and Optimization* 51.3 (2013): 2230–2260.
  - Damjanovic, Jovan, James M. Murphy, and Yu-Shan Lin. “Catboss: Cluster analysis of trajectories based on segment splitting.” *Journal of Chemical Information and Modeling* 61.10 (2021): 5066–5081.
  - Ramdas, Aaditya, Nicolás García Trillos, and Marco Cuturi. “On Wasserstein two-sample testing and related families of nonparametric tests.” *Entropy* 19.2 (2017): 47.

Acknowledgements

Some people and creatures I’d like to thank

Thank you so much for your support and kindness over the years:

My advisor James M. Murphy.

The members of my committee: Kasso Okoudjou, Abiy Tasissa, and Maria Cameron.

My colleagues at Analysis Group: Stefano Parravano, Upamanu Pathare, and Esteban Léon, and at the Naval Research Laboratory: Roger M. Oba, Laurie T. Fialkowski.

The wonderful weirdos of Sarah Lawrence College: Sarah Dennis, Olif Olson, Cole Cuthbert, Brooke Prakash, Isabella Romano, and Hannah Ford, as well as Profs. Mike Siff, Philip Ording, and Melissa Frazier.

The oddballs of Oregon: Thaís De Faria, Daniel Dougherty, Ben McDowell, Annie Price, Jess Lohrman, Adam and Stacy York, Luca Zocchi, Willow Davis, and the greater DWJ Lab family.

The many friends I’ve made at Tufts University: Merek Johnson, David Cohen, Brendan Mallery, Chandler Smith, Martin Buck, and Kecheng Li.

The inimitable Shane Frisco, you are an inspiration.

My family: my mother Linda especially, and also my brother Geof, my uncle Jack, and aunt Jeanne, and my in-laws Jeni, Kevin, Mason, Austin, and Riley Bates, for all their support over the years.

Finally, my amazing wife Dr. Hannah J. Bates, and our two perfect kittens Ezra and Kali, the best cats in this town or any other.

